Dynamic Methane Inventory for Oil and Gas Industry based on Artificial Intelligence

Jade Eva Guisiano

Sorbonne Université

Doctoral school: Computer Science, Telecommunications et Electronic

---

Dynamic Methane Inventory for Oil and Gas Industry based on Artificial Intelligence

---

Ph.D Thesis

Presented by

Jade Eva Guisiano

June 24th 2024

To obtain the doctor grade of Sorbonne Université
discipline: Computer Science

Jury:

Gabriele Facciolo	École Normale Supérieure	Reporter
Luis Guanter	Polytechnic University of Valencia	Reporter
Steven Wofsy	Harvard	Examiner
Silvia Liberata Ullo	Università degli Studi del Sannio	Examiner
Alexandre d'Aspremont	École Normale Supérieure	Examiner
Éric Moulines	École Polytechnique	Thesis co-director
Jérémie Sublime	ISEP	Thesis co-director
Thomas Lauvaux	Université de Reims	Thesis supervisor
Mark Radka	UN Environment Programme	Invited guest

Contents

List of Figures

List of Tables

Acknowledgements

I would first like to thank my thesis co-supervisors, **Jérémie Sublime** and **Eric Moulines** for their involvement, responsiveness and advice during these 3 years of thesis. A big thank you to Eric who allowed me to establish numerous scientific connections necessary for the completion of my thesis and a big thank you to Jérémie for his continued responsiveness and careful proofreading. I would particularly like to thank my co-supervisor **Thomas Lauvaux**, thanks to whom I was able to develop solid knowledge around methane during these 3 years. I am particularly grateful for his continued support which allowed me to complete my thesis successfully and in the best conditions. I also thank **Raja Chiky**, initial thesis director and friend, without whom my thesis subject would not have been possible. Thanks to **Patrick Gallinari** and **Emmanuel Gobet**, members of my monitoring committees, for their listening, availability and advice.

I would particularly like to thank the United Nations environmental program teams with whom I have been working for several years now: a big thank you to **Didier Salzmann** who helped initiate the adventure between the academic field and the UNEP and for his support during all these years. A huge thank you and deep gratitude to **Robert Rodriguez** for sharing his knowledge, his wise advice, and his precious help and kindness over many years. I would like to greatly thank **Mark Radka** who guided the choice of my thesis subject and who has always supported my work. I also thank the UNEP methane team, **Manfredi Caltagirone**, **Giulia Ferrini** and **Roland Kupers** who helped define and guide my thesis work. I would also like to thank **Jonathas De Mello**, with whom I had the chance to collaborate on the UNEP SDG-Meter project, who helped me on numerous occasions. Finally, I would like to thank all of the UNEP colleagues, including **Yann**, **Michel** and **Katy**, for their great sympathy and support.

I sincerely thank the Kayrros team including **Alexandre d'Aspremont**, **Alexis Groshenry**, **Clément Giron**, with whom I had the pleasure of collaborating and who have always been of great help.

I thank **Stefan Schwietzke**, **James Wang**, and **Daniel Zavala-Araiza** of the Environment Defense Fund who helped launch and initially guide my work.

I would like to thank **Zitely Tzompa-Sosa** for sharing her knowledge and with whom I had the pleasure of collaborating on the writing of my first journal article. A big thank you to **Evan Sherwin** who has always been very helpful, for sharing this knowledge, this advice and sharing his network. I also thank **Balaji Padmanabhan** for his advice, his support and our rich discussions around Artificial Intelligence.

I thank my fellow professors including **Patricia Condes-Cespedes**, **Maria Trocan**, **Hélène Urien** and **Zakia Kazi** who gave me the opportunity to teach various modules for the ISEP engineering cycle for several years, I also thank them for their sympathy and support. I would also like to thank my colleagues and friends, doctors and doctoral students, with whom I had a great time: **Maurras Togbe, Shufan Jiang, Mariam Barry, Guillaume Lachaud, Arthur Vervaet, Nan Ding, Abir Aissa, Seoyoung Oh, Abdul Qadir Khan** ...

I would also like to thank the members of **Climate Change AI (CCAI)** of which I am a volunteer, who allowed me to have access to rich exchanges and experiences at the intersection of my two areas of interest.

I sincerely thank my friends for their presence and support who have given me a lot in many aspects: **Ouafa, Domenico, Roberto, the Montoya family and the Garcia-Pervan family**. I also thank all my friends and acquaintances from Changsha (China) including **Pan Yaolin** and **Zhang Xinyu** who were there for me at a crucial moment and who have greatly contributed to my progress and personal development.

Finally, I thank my **Guisiano** and **Andrieu** families for their endless support and presence.

Abstract

Atmospheric methane is responsible for roughly half of the global warming since pre-industrial times, relative to the net total human influence. The oil and gas sector ranks second among anthropogenic methane sources despite the possibility of a 39% reduction of its emissions at no net cost. However, the lack of reliable emissions data impairs the ability of governments to implement effective mitigation actions at the scale and speed needed to achieve the objectives set by the Global Methane Pledge. In order to define methane-specific targets, policies, and strategies, countries are used to rely on national methane emissions inventories. These last ones, mostly derived using bottom-up methods (emissions factors) are well-known in literature to demonstrated systematic underestimation of methane emissions from the oil and gas (O&G) supply chain. This calls into question the use of bottom-up methodologies to estimate emission inventories, which are then used to design regulatory guidelines for methane emissions mitigation. Incomplete and non accurate information about actual emission levels is a key barrier to reducing methane emissions. However, a growing amount of methane emissions data emerge through the regular launch of new point source satellites dedicated to methane concentration measurements with a higher resolution, greater coverage, and have more sensitive detection thresholds. With adequate processing techniques, the data issued from these satellites monitoring technologies is a key component to make possible the characterization of the level and nature of methane emissions. This thesis works then proposes an end-to-end framework for a novel dynamic inventory based on the use of artificial intelligence methods. This framework allows to automatically dress methane emissions profiles by oil and gas infrastructures, site and operators. The proposed framework is divided into two components :

1. Automated detection and recognition of oil and gas infrastructures with the use of object detection algorithms;

2. Automated association of point source methane detections to concerned oil and gas infrastructures, sites and operators with the use of a clustering based method.

Our framework, when launched repeatedly over time, enables the automated constitution of multi-level emissions profiles (time series). These emissions profiles, acquired on a long term period will allow the characterization of emitting behaviors and then could be used at the base of methane mitigation regulation decisions.

Keywords: Methane; Oil & Gas; GHG inventory; Deep learning; Remote sensing.

Chapter 1

Introduction

Contents

1.1 Context

Climate change is at the origin of the alteration of Earth's climate patterns, including increases in global average temperatures and shifts in precipitation patterns. These changes have far-reaching impacts on ecosystems, biodiversity, and human societies, leading to disruptions in agriculture, water resources, and infrastructure, as well as exacerbating extreme weather events such as hurricanes, droughts, and wildfires. Global warming is a key component of climate change and is primarily driven by the accumulation of greenhouse gases in the Earth's atmosphere, such as carbon dioxide and methane, due to human activities like burning fossil fuels and deforestation. These gases trap heat from the sun, causing the Earth's surface temperature to rise, leading to various climatic alterations including melting polar ice caps, rising sea levels, and changes in precipitation patterns. Methane is re-

sponsible for 50% [43] of the global warming since pre-industrial times. By definition, methane is a critical short-lived climate pollutant with a global warming potential over 80 times that of CO2 on a 20-year timescale.The oil and gas (O&G) sector is the second largest anthropogenic methane emission source. In contrast to agriculture and wetlands, the O&G industry has a long history of reducing methane emissions due to safety concerns, and methane can be sold as natural gas, so reducing emissions has an important financial benefit. The International Energy Agency (IEA) estimates that the industry can reduce its global emissions by 78% and that up to 39% [41] of these reductions can be achieved without a financial compromise. Efficient methane abatement strategies directly depend on the accurate characterization of emission sources. However, the lack of reliable emissions data has made it hard for governments to carry out targeted action at the scale and speed needed to achieve the objectives of the Global Methane Pledge (GMP).

1.2 Motivations

In order to define methane-specific targets, policies, and strategies, countries rely on national methane emissions inventories. These last ones are mostly derived using bottom-up methods which estimates overall methane emissions by using activity data (e.g., the number of facilities or the extent of operations) multiplied by standardised emission factors which are the average amount of methane emitted per facilities or operations. However, various studies (*e.g.* [1, 11, 93]) have demonstrated the systematic underestimation of methane emissions from the O&G supply chain. This finding calls into question the use of bottom-up methodologies to estimate emission inventories, which are then used to design regulatory guidelines for methane emissions reduction. Incomplete and non accurate information about actual emission levels is a key barrier to reducing methane emissions. However, a growing amount of methane emissions data emerge through the regular launch of new satellites dedicated to methane concentration measurements with a higher resolution, greater coverage, and have more sensitive detection thresholds. Technological innovation around space-borne methane measurement tools tends to focus its long-term vision in the area of quasi-continuous monitoring. These advances in satellites monitoring technologies and their processing techniques are a key component to permit the characterization of the level and nature of methane emissions through a quasi-continuous monitoring. By combining the respective advantages

of top-down and bottom-up methane emission estimations will ultimately improve their precision.

The primary objective of the International Methane Emissions Observatory (IMEO), established under the auspices of the United Nations Environment Programme (UNEP), is to furnish near-real-time data pertaining to methane emissions originating from the fossil fuel sector. This initiative involves the assimilation of data from diverse sources, including satellite platforms such as TROPOMI, GHGSat, and MethaneSAT. Effectively harnessing these datasets to inform methane emissions reduction policies necessitates the establishment of a robust linkage between emissions data and their respective sources, thereby enabling the characterization of emitting behaviors. Achieving this entails attributing all methane detections to the specific oil and gas infrastructures, sites, and operators from which they originate. Given the disparate nature of methane emission detections emanating from various satellites and locations, the attribution process necessitates automation through the development of a new dedicated,framework based on artificial intelligence methodologies.

Indeed, in this context, artificial intelligence methods will assume a central role in automating and optimizing all steps of this process simultaneously. Such a framework will facilitate the expansion of our understanding of the spatial distribution, occurrence, and characterization of methane emissions.

1.3 Contributions

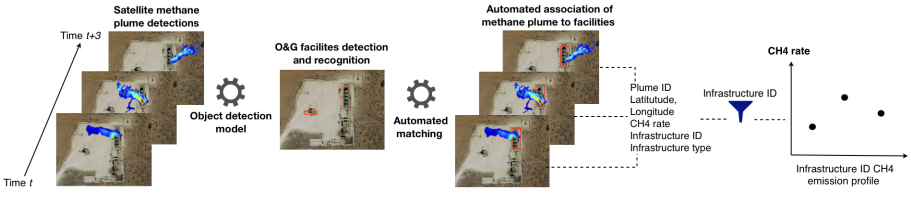


Figure 1.1: End-to-end dynamic methane emissions inventory framework. *Images source : @Google earth.*

The thesis contributions are the following:

- Conception of a benchmark dataset of high resolution satellite images of oil and gas infrastructures;
- Automated oil and gas infrastructure detection and recognizing based on object detection algorithms;
- Determination of object detection algorithm's pre-training effects on their performances;
- Sensitivity and robustness of object detection models to satellite based adversarial attacks;
- Methane Emissions automated association to oil and gas supply-chain sites, oil and gas operators and ground surveys extension;
- End-to-end framework for Automated determination of oil and gas infrastructures emissions profile illustrated by the Figure 1.1

1.4 Thesis Outline

The thesis manuscript is organized as follows :

- Chapter 2 introduces the context of methane emissions in the oil and gas industry and its role in global warming, the existing tools and methods for measuring methane emissions as well as the current regulation framework and its challenges.
- Chapter 3 first presents the state of the art of methane emissions monitoring methods from automated methane plume detection to methane emissions forecasting based on the use of artificial intelligence. Then, presents the performance of 3 families of object detection algorithms for the automatic recognition of oil and gas infrastructures on high-resolution satellite images. It also presents the effect of pre-training and satellite image noises on object detection performances.
- Chapter 4 presents a method for automatically associating methane emissions with different parts of the oil and gas supply chain and operators, based on the use of a clustering algorithm. It also describes a framework for the automatic association of methane plumes with oil and gas infrastructures.

- Chapter 5 first presents a end to end framework for a dynamic and intelligent methane emissions inventory. Then, it discusses its potential contributions to methane mitigation policies design.

- Chapter 6 concludes the thesis works and introduces perspectives of futures works.

1.5 Publications

Journal

- Jade E Guisiano, Zitely A. Tzompa-Sosa, Thomas Lauvaux, Éric Moulines, Jérémie Sublime. "Dynamic & Intelligent Methane Emissions Inventory (DIMEI) Framework: Next-generation methane emission inventory for oil and gas industry based on Artificial Intelligence" Iscience CellPress (2024) [peer-review]

Articles

- Jade E Guisiano, Domenico Barretta, Éric Moulines, Thomas Lauvaux, Jérémie Sublime. "Object detection models sensitivity & robustness to satellite-based adversarial attacks". International Geoscience and Remote Sensing Symposium IGARSS (2024), Athens, Greece. [peer-review]

- Jade E Guisiano, Thomas Lauvaux, Éric Moulines, Jérémie Sublime "Oil and Gas Automatic Infrastructure Mapping: Leveraging High-Resolution Satellite Imagery Through Fine-Tuning of Object Detection Models". International Conference On Neural Information Processing ICONIP (2023), Changsha, China.

- Jade E Guisiano, Thomas Lauvaux, Claudio Cifarelli, Éric Moulines, Jérémie Sublime. "O&GProfile : An automated method for attribution of satellite methane emissions detections to oil and gas sites and operators". International Conference on Machine Learning and Data Mining MLDM (2023), New-York, United States.

Workshop

- Jade E Guisiano, Zitely A. Tzompa-Sosa, Thomas Lauvaux. "Artificial intelligence for dynamic and intelligent methane inventory". European Geosciences Union EGU General Assembly (2024), Vienna, Austria.
- Jade E Guisiano, Éric Moulines, Thomas Lauvaux, Jérémie Sublime. "Artificial Intelligence for Methane Mitigation : Through an Automated Determination of Oil and Gas Methane Emissions Profiles". Conference on Neural Information Processing Systems NeurIPS,(2023), Nouvelle-Orléans, Louisiane, United States.
- Alba Lorente, Daniel Zavala-Araiza, Daniel J. Varon, Lu Shen, Mengyao Liu, Yuzhong Zhang, J. Pepijn Veefkind et al. "Comprehensive Synthesis Of TROPOMI-Based Methane Emissions From The Permian Basin Estimated With State-Of-Art Methodologies." American Geophysical Union AGU (2023), Washington, DC, United States.
- Jade E Guisiano, Thomas Lauvaux, Claudio Cifarelli, Éric Moulines, Jérémie Sublime. "O&GProfile : Automated attribution of GHGSat point source methane emissions detections to O&G infrastructures for site emissions profile analysis (Permian)". European Geosciences Union EGU General Assembly (2023), Vienna, Austria.
- Alba Lorente, Daniel Zavala-Araiza, Daniel J. Varon, Lu Shen, Mengyao Liu, Yuzhong Zhang, J. Pepijn Veefkind et al. "Comprehensive synthesis of TROPOMI-based methane emissions from the Permian basin estimated with state-of-art methodologies." International Workshop On Greenhouse Gas Measurements From Space IWGGMS (2023), Paris, France.

Chapter 2

Methane Emissions from the Oil and Gas industry

This chapter outlines the definitions and background of methane, the oil and gas industry, and their contributions to global warming. It further discusses existing regulations for methane mitigation, their constraints, and potential solutions, introducing our research.

Section 2.1 defines methane and its sources, highlighting its role in global warming. Section 2.2 presents tools for methane measurement and associated quantification methods. Section 2.3 delineates the fossil fuel sector, with a focus on the oil and gas industry's specificities concerning methane emissions. Lastly, Section 2.4 reviews current methane mitigation policies, their limitations, and introduces the foundational aspects of this thesis, aiming to offer efficient methane mitigation solutions.

Contents

2.1 Methane and Global Warming

2.1.1 Methane and its impacts

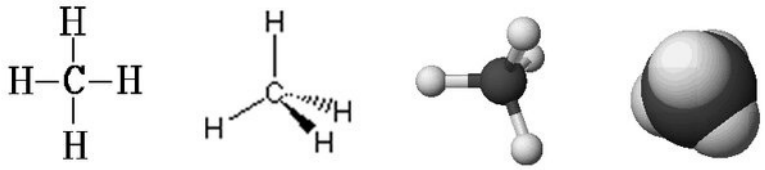


Figure 2.1: Different representations of methane (CH_4) molecule [118].

Methane is a chemical compound with the molecular formula CH_4, composed of hydrogen (H) and carbon (C) as depicted in Figure 2.1. It is an odorless, colorless greenhouse gas (GHG) lighter than air. Following carbon dioxide, methane ranks as the second most significant contributor to climate change and also impacts human and environmental health as a pollutant[73]. While methane's atmospheric lifetime is shorter than that of CO_2, its warming potential is substantially higher. The Intergovernmental Panel on Climate Change (IPCC) estimates that the climate impact of one unit mass of methane is 84 times that of CO_2 over a 20-year period and remains 28 times greater over 100 years. These attributes position methane as a promising target for effective GHG emissions reduction.

The greenhouse effect[60, 5, 61] is a thermal phenomenon in which solar radiation reaching Earth's atmosphere is partly reflected back into space by the atmosphere, white clouds, and bright surfaces, such as polar ice caps. Unreflected incident rays are absorbed by the Earth's atmosphere and/or surface, providing heat (energy) that is subsequently radiated back as infrared rays (black body radiation). A portion of this radiation is absorbed

by greenhouse gases (GHGs). In a subsequent stage, the absorbed heat is re-emitted in all directions, including back towards Earth, creating the greenhouse effect that elevates surface temperatures. This mechanism maintains a positive temperature balance on Earth; without it, temperatures would be considerably lower. However, increasing GHG concentrations intensify this effect, trapping more radiation and leading to rising surface temperatures. Over the past century, global average temperatures have increased by approximately 1.25°C (2.25°F), with GHG increases playing a significant role [39][141].

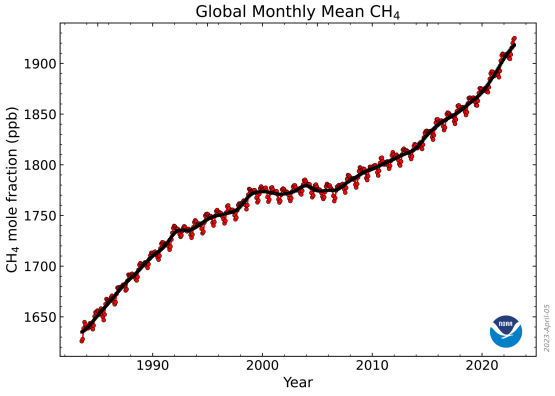


Figure 2.2: Globally-averaged, monthly mean atmospheric methane abundance determined from 1983 to 2022. *Image source: National Oceanic and Atmospheric Administration (NOAA).*

As shown in Figure 2.2, methane concentrations in the atmosphere surpassed 1,900 parts per billion (ppb) last year, nearly tripling pre-industrial levels. In 2022, methane concentration rose by 14.0 ppb, representing the fourth-largest annual increase since the National Oceanic and Atmospheric Administration (NOAA) began systematic measurements in 1983. This rise continues the trend observed in 2020 and 2021. The Global Methane Assessment (GMA) 2021 reported a notable acceleration in atmospheric methane concentrations during the 2010s, resulting in five-year average growth rates not observed since the 1980s.

The observed methane levels substantially exceeded those projected in the 2°C scenario outlined in the IPCC 2013 Assessment[42]. Since the publication of the Global Methane Assessment (GMA) in 2021, official atmospheric values for 2020 have been established by the World Meteorological

Organization (WMO) Global Atmosphere Watch Programme (GAW). The globally averaged surface methane level for 2020 was measured at 1889 ± 2 ppb, indicating a 262% increase from pre-industrial levels. The annual increase from 2019 to 2020 was 11 ppb, surpassing the average annual growth rate observed over the previous decade. Data analysis from U.S. National Oceanic and Atmospheric Administration (NOAA) marine boundary layer sites, accounting for approximately 40% of the GAW network, identified a methane increase of 17.0 ppb in 2021, marking the highest annual increase in the 38-year record.

Preliminary analysis of methane growth rates, derived from satellite total column data using SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT products from SRON, revealed a rate of 16.3 ppb per year in 2021. This rate represents the highest value observed in the 2003-2021 record. Both ground-based and satellite observations indicate that atmospheric methane levels are continuing to rise rapidly as the world progresses into the 2020s [122, 22, 120].

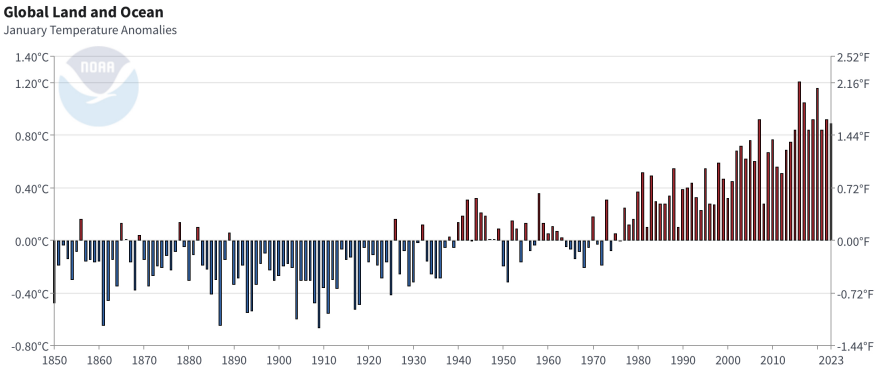


Figure 2.3: Global temperature anomaly by year in Fahrenheit and degree Celsius from 1880 to 2022. *image source: National Oceanic and Atmospheric Administration (NOAA).*

Methane contributes to approximately 50% [43] of global warming since pre-industrial times, relative to the net total human influence. Global warming refers to the long-term increase in Earth's average surface temperature due to the greenhouse effect. While Earth's surface warming has been non-uniform, the overall trend in globally averaged temperature indicates more regions experiencing warming than cooling. According to NOAA's 2023 Annual Climate Report[79] and Figure 2.3, the combined land and ocean

temperature has increased at an average rate of 0.11°F (0.06°C) per decade since 1850, resulting in a total increase of approximately 2°F. The rate of warming has accelerated since 1982, with an average increase of 0.36°F (0.20°C) per decade. The latest Synthesis Report from the IPCC[43] attributes this warming trend to human activities, particularly GHG emissions, with global surface temperatures reaching 1.1°C above pre-industrial levels (1850-1900) between 2011 and 2020[122]. If GHG concentrations and associated global temperatures continue to rise, widespread impacts on global climate are anticipated, with some already underway [13] [123]. These impacts are expected to have significant implications for global food production, human health, and the habitability of coastal and arid regions.

2.1.2 Methane sources

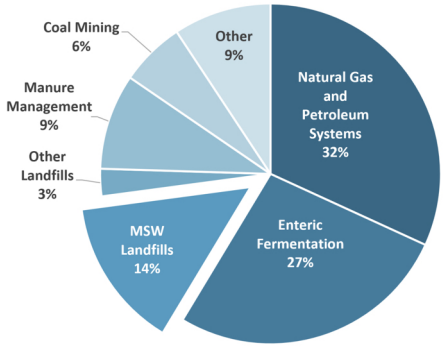


Figure 2.4: 2019 Anthropogenic Methane emission by sources in United States. *Image source: United States Environmental Protection Agency (EPA).*

Methane emissions can be categorized into two groups: natural sources and anthropogenic sources. According to the EPA, natural sources contribute to approximately 40% of global methane emissions. Among these, natural wetlands are the predominant source, emitting methane through bacterial decomposition of organic materials in oxygen-deprived conditions. Additional minor sources include termites, oceans, sediments, volcanoes, and wildfires. Anthropogenic sources, as shown in Figure 2.4, account for 60% of global methane emissions. Thus, over half of all methane emissions originate from the following primary human activities [114]:

- **Agriculture**: Agriculture represents the largest anthropogenic source of methane emissions, accounting for approximately 41% of total anthropogenic emissions. This is primarily attributed to enteric fermentation in ruminant livestock, which generates methane as a digestion byproduct. Additionally, manure management and rice cultivation contribute to methane emissions from this sector.
- **Energy Production**: The energy sector, encompassing the extraction, production, and utilization of fossil fuels, is a significant contributor to methane emissions, representing around 35% of the total. This includes emissions from oil and natural gas production, coal mining, as well as the combustion of fossil fuels for electricity generation, heating, and transportation.
- **Waste Management**: Organic waste management practices, such as landfilling, composting, and wastewater treatment, can produce methane through anaerobic decomposition. Landfills stand out as a significant source of methane emissions, contributing approximately 20% of total anthropogenic emissions.
- **Others**: Several other categories also contribute to methane emissions, albeit with lower contributions. These include specific industrial processes and biomass burning, encompassing wildfires, deforestation, and peatlands.

These emissions sources, by releasing methane, contribute to global warming, which in turn triggers various mechanisms leading to the release of greenhouse gases (GHGs):

- **Methane Hydrates**: Methane hydrates are ice-like structures containing methane molecules trapped within water molecules. These deposits, found in permafrost regions and deep ocean sediments, have the potential to release large amounts of methane if they melt or are disturbed.
- **Permafrost**: Permafrost refers to soil that remains frozen throughout the year and contains substantial amounts of organic carbon. As permafrost thaws due to climate change, organic matter decomposes, releasing methane and carbon dioxide into the atmosphere.
- **Methane Seeps**: Methane seeps are areas where methane gas escapes from the seafloor into the water column. These seeps can occur naturally or as a result of human activities, such as oil and gas drilling.

In summary, methane emissions originate from a range of natural and human-made sources, with agriculture, energy production, and waste management serving as the largest anthropogenic contributors.

2.2 Methane emissions measurement

2.2.1 Measurement instruments

The concentration of methane in the atmosphere is determined by its emissions. Elevated methane emissions result in increased atmospheric concentrations, which are typically measured in parts per million (ppm) or parts per billion (ppb). The estimation and quantification of methane emissions utilize methane-sensitive sensors employing various specific techniques (cf. Section 2.2.2). These sensors facilitate the calculation of emission rates, distinguishing between source emissions and background levels. The design and type of sensor dictate the measurable parameters and operating conditions. Sensor placement can be categorized into two primary approaches: in-situ measurements and remote sensing.

In-situ sensors necessitate direct contact with methane molecules within a plume for chemical or physical interactions and, therefore, must be situated on-site, either near or directly within the methane plume [119][33]. Conversely, other sensors, suitable for both in-situ and remote sensing[150] applications, operate based on the detection of electromagnetic signals, such as infrared light from the sun or a laser. As electromagnetic radiation traverses the atmosphere, specific wavelengths are absorbed by various molecules, each exhibiting a unique absorption signature. Multiple detection instruments, including optical gas imaging and multi-spectral/hyperspectral imaging, leverage these methane absorption characteristics for detection and measurement.

Methane absorbs infrared radiation across a specific range of wavelengths, with pronounced absorption in certain wavelength regions. Methane demonstrates absorption characteristics throughout the infrared spectral range of 0.75-14 μm. Additionally, other atmospheric gases, including CO2, CO, O3, N2O, and water vapor, possess unique absorption signatures that partially overlap with methane absorption wavelengths. Consequently, instrument systems must be meticulously designed to differentiate and isolate methane

signals from potential interference caused by these other gases.

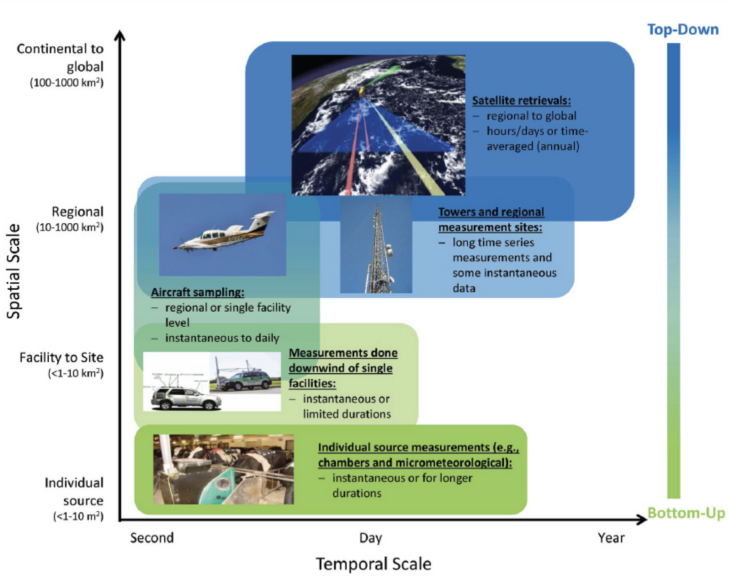


Figure 2.5: Tools for Methane emissions measurement and its characteristics

[77].

Measurements of methane concentrations can be conducted across various spatial and temporal scales, ranging from broad global assessments to localized evaluations of individual sources. The temporal interval for concentration measurements can vary from long-term, annual assessments to short-term, near-real-time evaluations. These parameters are influenced by the sensor characteristics and their placement. Sensor locations can vary from ground-based installations to satellites, enabling measurements at different spatial scales based on altitude.

As depicted in Figure 2.5, at high altitudes (100-10,000 km from the emission source), satellite-based sensors can cover extensive spatial scales, including global, continental, and regional levels (100-1000 km^2). The measurement frequency can range from continuous to hourly, daily, or annual averages. At intermediate altitudes (100m-10km), aircraft campaigns and tower-based measurements enable regional-level emissions readings (10-100 km^2) with instantaneous or daily measurements. Additionally, targeted measurements can be obtained for specific sites ($<$1-10 km^2), such as indus-

trial facilities, using aircraft campaigns, low-altitude drones (10m-10km), providing instantaneous to daily measurements. Lastly, as previously mentioned, in-situ sensors can be positioned near or directly within the emission source, providing instantaneous or longer-term measurements.

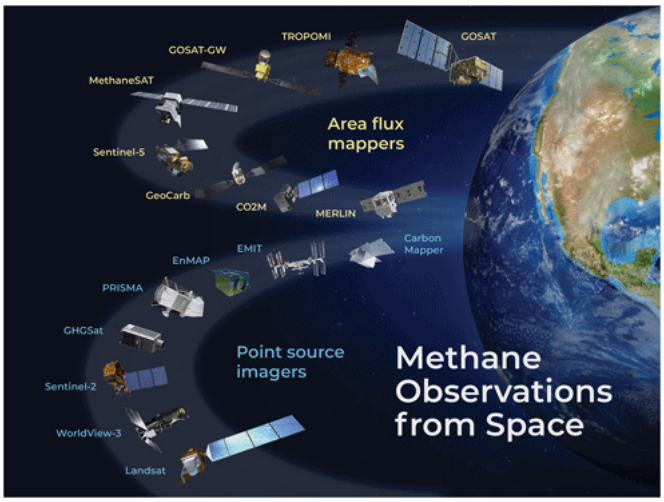


Figure 2.6: Satellite instruments for observation of methane in the shortwave infrared (SWIR) by area flux mapper and point sources imagers.

The current satellite landscape includes at least 16 satellites dedicated to monitoring methane concentrations, as shown in Figure 2.6. Each satellite possesses distinct characteristics:

- **Spectral resolution**: Refers to the capability of a satellite sensor to measure specific wavelengths within the electromagnetic spectrum. A finer spectral resolution corresponds to a narrower wavelength range for a particular channel or band.
- **Spatial resolution**: Represents the smallest object (pixel) that the sensor can resolve or the ground area imaged. Each pixel corresponds to a specific area of the Earth's surface and has an associated intensity value and location address within the two-dimensional image.
- **Revisit time**: Denotes the interval between two successive observations of the same point on Earth by the satellite. This parameter is influenced by the satellite's technical capabilities, such as attitude control and the electronic steering capability of the payload antenna.

Satellite instruments can be categorized into two types: area flux mappers and point source imagers, each with distinct characteristics. Area flux mappers are characterized by high precision (<1%) and pixel sizes ranging from 0.1 to 10 kilometers, facilitating the quantification of total methane emissions at regional to global scales. These instruments are particularly suited for assessing area sources, which may encompass a vast number of individually small emitters that collectively contribute to significant emissions. Notable examples of area flux mappers include GOSAT and TROPOMI, which offer continuous daily mapping and long-term methane trend analysis, respectively.

Conversely, point source imagers, such as the GHGSat constellation and various hyperspectral and multispectral land imaging sensors (e.g., PRISMA, Sentinel-2, Landsat-8/9, WorldView-3), feature finer pixel sizes (<60 meters). They are designed to detect individual point sources by imaging their plumes, as illustrated in Figure 2.7. These imagers have detection thresholds ranging from 100–10,000 kilograms per hour, enabling the monitoring of large point sources[124].

Figure 2.7: Methane plume enhancement over Libya by Satellites TROPOMI Sentinel 5P and GHGSat *Image source: European Spatial Agency.*

The constellation of satellites is planned to be extended through the continuous development of technologies that offer improved spectral resolution, spatial resolution, and temporal coverage, ultimately enabling quasi-real-time monitoring.

2.2.2 Emissions quantification

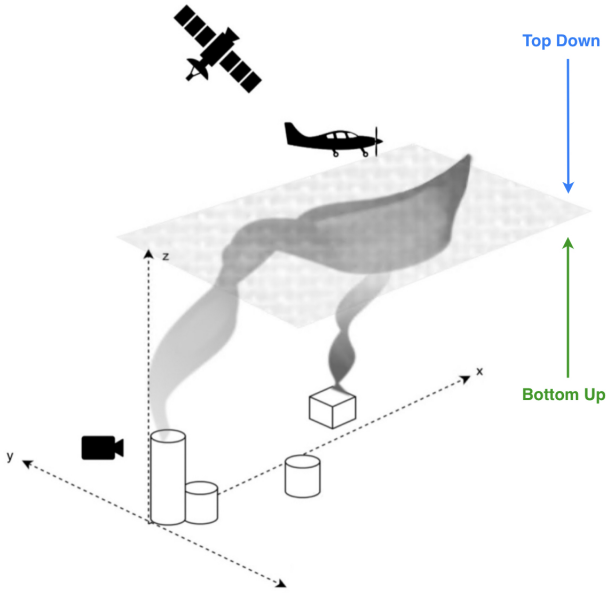


Figure 2.8: Top down & Bottom-up perspectives of methane emissions estimation.

Measurements of methane concentrations obtained at various altitudes require specific transformations to estimate methane emissions. These transformations are influenced by both the sensor's position and the selected study area. The methodologies for estimating emissions are categorized into two approaches: bottom-up and top-down, as illustrated in Figure 2.8:

Bottom-up. methods estimate emissions using activity data (e.g., industry segment, annual production, type of component or equipment) and emission factors tailored to the specific facility or component analyzed. Formally, bottom-up methods utilize activity information, such as the number of facilities or the scale of operations, which is then multiplied by standardized emission factors, including default values or leak rates specific to certain equipment types.

According to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories [81], emission factors vary based on the level of methodological

complexity employed in these engineering calculations. The IPCC identifies three methodological tiers: the first corresponds to the basic method, requiring minimal information from the emission source. As the tier number increases, more information is needed, leading to more complex calculations. Regardless of the tier used, emission factors are generic for the specific industry segment, facility, or component to which they apply. In more detailed cases, these factors are derived from direct measurements at the emission source, which are then extrapolated to represent the entire population of similar emission sources at regional or national levels. This approach introduces inherent uncertainty into the calculations.

For bottom-up emissions estimates, models serve multiple purposes. They are utilized to convert methane concentration data into methane emissions estimates and to translate methane measurements taken away from a source into emissions measurements using atmospheric dispersion models. Subsequently, these models are employed to convert emissions measurements across a sampled population into comprehensive emissions estimates at the facility, regional, national, or global level using statistical models.

Top-down. Rather than estimating emissions from a representative sample of devices, as done in the bottom-up method, the top-down method relies on observations of methane concentration in the atmosphere, obtained using sensors located at high altitudes. To determine the location and quantify the magnitude of a methane emission source, the transport of methane from the emission source to the measurement location must be simulated. Subsequently, methane concentrations are converted to an emission flux using an atmospheric inversion model. This model relies on atmospheric methane measurements and an atmospheric transport model to infer the most likely distribution of emissions or methane fluxes at various levels of the Earth's atmosphere. To calculate flux rates, atmospheric transport models incorporate factors such as wind, atmospheric conditions, and background methane concentrations. Inversion models are categorized based on the targeted scale of estimation:

- **Global-regional inversion**: Quantifying emissions on global-regional scales, with multiple contributing sources, necessitates a broader approach. An ensemble of atmospheric observations from area flux mapper instruments is fitted to a 2D field of emissions through inversion of a 3D chemical transport model (CTM) that correlates emissions with atmospheric concentrations [128]. The process involves simulating

the movement of methane molecules, utilizing observed emissions and prior data (bottom-up inventory), to ascertain their source. Through optimization, the model iteratively refines prior emission estimates until a consensus between simulated and observed concentrations is achieved. Two main types of optimization models can be employed:

  - **Eulerian**: These models concentrate on fixed locations and offer a comprehensive, continuous, mass-preserving representation of the atmosphere. They can also be retroactively integrated to derive source fingerprints. Eulerian models are primarily used for inverse analyses of methane observations at the satellite level, where a large number of receiving points are present [29][82][115].
  - **Lagrangian**: These models simulate the movement of air parcels and the alterations they undergo due to their motion. The primary assumption of Lagrangian models is the absence of horizontal dispersion of pollutants, implying that once materials enter the column, they are not eliminated by mixing and dilution with the surrounding air. Lagrangian models are directly integrable backwards in time, allowing for cost-effective calculation of the source footprint contributing to concentrations at a specific receptor point [130][3][83][85].

  The combination of Eulerian and Lagrangian models is also feasible [94, 6, 55], providing advantages such as high-resolution simulations near measurement sites, minimized errors due to the aggregation of emissions, and the flexibility to employ any combination of suitable models without code modification. This approach allows for sensitivity to the entire emissions field and facilitates estimation using traditional statistical methods without multiple inversion steps.

- **Point-source inversion**: Targeted atmospheric measurements of methane can quantify emissions on smaller scales (point source, urban area, oil and gas basin). The behavior of methane plumes is influenced by atmospheric turbulence, advection, and the intensity of the emission source. Various methods are available to estimate the source rates of these plumes, including Gaussian plume inversion[102][103], local mass balance[9], Gauss theorem[12], cross-sectional flux method (CSF) [99, 125, 126, 96], and the Integrated Mass Enhancement (IME) [125, 78, 126, 56] method. The optimal approach depends on the availability of meteorological information and the instrument's ability to map the observed plume structure, which is influenced by pixel size, measurement noise, the capacity to define the local background, and flow complexity, including wind shear effects[128].

> Given the instantaneous nature and variability of plumes in downwind transport, the IME and cross-sectional methods are deemed the most suitable for estimating source rates. Both methods require the determination of local or effective wind speed, with the IME method demonstrating lower error rates in effective wind speed determination. Both methods are extensively applied to retrieve point source rates from satellite observations, yielding consistent results[127].

Top-down and bottom-up approaches each offer distinct advantages and face specific limitations. The bottom-up approach enables the direct quantification of emissions from known sources or facilities, attributing emissions to specific sources. However, this method necessitates comprehensive accounting of activity data, precise emission factors, and a representative sample size to characterize the scaled population accurately. The vast number of sources, which can vary substantially over time and space, poses a significant challenge for establishing precise regional or national emissions inventories. Moreover, full access to all sources is not always feasible.

Conversely, top-down approaches encompass contributions from all sources, including unknown or underestimated sources. However, attributing emissions to specific processes or activities is challenging due to the intermingling of sources and limited detection capabilities of individual sources or installations.

While bottom-up approaches offer valuable insights into the specific sources of emissions and suggest targeted actions for emission reduction, top-down approaches can reveal unexpected and often substantial leaks that may remain unidentified using a bottom-up approach.

2.3 Oil and Gas Industry specificities

2.3.1 Fossil fuel sector

The fossil fuel sector encompasses the industry engaged in the extraction, processing, and distribution of non-renewable energy sources, including bioenergy, coal, oil, and natural gas. It spans activities from exploration and drilling to refining and transportation, playing a pivotal role in global energy production and consumption.

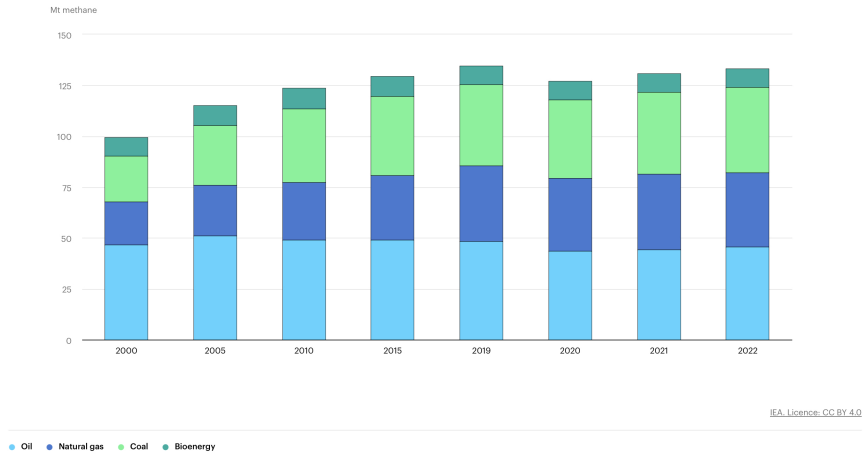


Figure 2.9: Methane emission in (Mt) from energy sector between 2000 and 2020. *Image source: International Energy Agency (IEA).*

According to the International Energy Agency (IEA)[32], the global energy sector contributed an estimated 135 million tonnes of methane emissions in 2022 (cf. Figure 2.9), marking a slight increase from the previous year. Methane emissions from the energy sector constitute nearly 40% of the total methane emissions attributed to human activities, ranking second only to agriculture (cf. Section 2.1.2). Coal, oil, and natural gas operations individually contributed approximately 40 million tonnes of methane emissions, with an additional nearly 5 million tonnes originating from leaks in end-use equipment. Approximately 10 million tonnes of emissions were attributed to the incomplete combustion of bioenergy, primarily from the traditional use of biomass. As illustrated in Figure 2.9, methane emissions from natural gas activities were estimated at 36.7 million tonnes, while those from oil activities were estimated at 45.6 million tonnes in 2022, representing approximately 62% of the global energy sector emissions.

Methane emissions from the fossil fuel sector significantly contribute to global greenhouse gas emissions, with certain countries playing a more substantial role in this context. As depicted in Figure 2.10, the leading methane-emitting countries from fossil fuel activities include China, Russia, the United States, Turkmenistan, and Iraq. China, the world's largest coal producer and consumer, emits considerable methane from coal mining activities. Similarly, Iraq, a major oil and gas producer, significantly

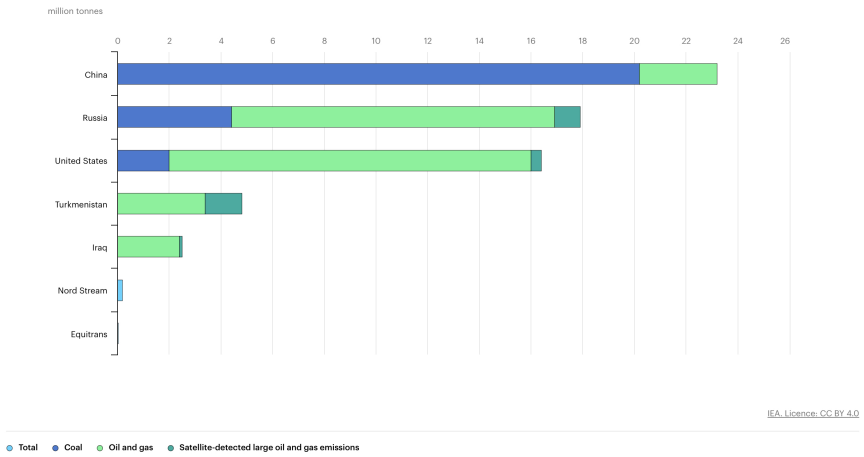


Figure 2.10: Large leaks from fossil fuel operations in 2022 by country. *Image source: International Energy Agency (IEA).*

contributes to methane emissions from fossil fuel operations. The United States accounts for a significant share of methane emissions from oil and gas activities, being the largest global producer in these sectors. Russia, another major oil and gas producer, also contributes significantly to methane emissions from its fossil fuel sector. Collectively, these countries account for a substantial portion of global methane emissions from the fossil fuel sector.

Despite these challenges, there is considerable potential to reduce methane emissions from the energy sector. It is estimated that approximately 70% of methane emissions from fossil fuel operations could be mitigated using existing technologies. The oil and gas (O&G) sector, being the second largest anthropogenic methane emission source, can achieve emission reductions through the adoption of existing and proven technologies[32]. Moreover, unlike other industries, the O&G sector has a longstanding history of reducing methane emissions due to safety concerns, and methane can be marketed as natural gas, providing a significant financial incentive for emission reduction efforts. The International Energy Agency (IEA) estimates that the industry can reduce its global emissions by 78%, with up to 39% of these reductions achievable without financial compromise[41].

2.3.2 Oil and Gas industry presentation

The oil and gas industry, also referred to as the petroleum industry, encompasses the exploration, extraction, production, refining, and distribution of petroleum and natural gas resources. These primarily hydrocarbon-based resources are crucial for fulfilling global energy demand, supplying fuel for transportation, electricity generation, and various industrial processes.

The industry operates within a multifaceted network of companies, including multinational corporations, national oil companies, and independent operators, operating across diverse regions and countries. While the oil and gas sector plays an essential role in global economic and geopolitical dynamics, it also confronts challenges related to environmental sustainability, fluctuating oil prices, and the transition to renewable energy sources.

Globally, there are over 25,000 oil and gas basins of varying sizes. The world's largest oil fields are predominantly located in the Middle East, Brazil, Mexico, Venezuela, Kazakhstan, and Russia. However, the most significant oil and gas basins are those that have historically yielded substantial quantities of hydrocarbons and continue to be primary contributors to the global energy supply. Some of the most notable basins include:

- **The Permian Basin**: Situated in West Texas and southeastern New Mexico, the Permian Basin stands as one of the largest and most productive oil and gas basins globally. It has consistently contributed to the United States' oil and gas output for decades, drawing substantial investment.
- **The Ghawar Field**: Located in Saudi Arabia, the Ghawar Field represents the largest conventional oil field globally and has served as a significant oil source for the international market over several decades.
- **The Marcellus Shale**: Positioned in the northeastern United States, the Marcellus Shale ranks among the largest natural gas-producing regions in the country. Recent years have witnessed notable production growth, attributable to advancements in hydraulic fracturing technology.

O&G activities is divided in 3 sectors, and each sector has distinct sources of methane emissions :

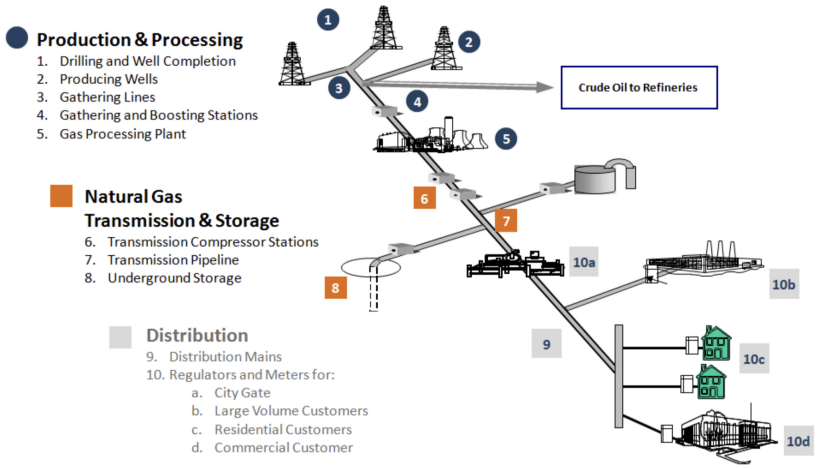


Figure 2.11: Oil and Gas Supply-chain. Upstream: (1,2), Midstream: (6,7,8,9), Downstream: (3,4,5) *Image source: U.S Environmental Protection Agency (EPA).*

- **Upstream**: The upstream sector encompasses the initial stages of the oil and gas industry, focusing on locating and extracting hydrocarbon resources from the earth's subsurface. This includes exploration, drilling, and production of crude oil and natural gas. Methane emissions in the upstream sector primarily originate from wellheads, flares, pipelines, and storage tanks;
- **Midstream**: The midstream phase encompasses the transportation, storage, and processing of crude oil and natural gas subsequent to their extraction. It also includes the conveyance of these resources from production sites to refineries and other downstream facilities. Methane emissions in the midstream sector primarily arise from pipelines, compressor stations, and storage facilities.
- **Downstream**: The downstream phase encompasses the conversion of crude oil into diverse petroleum products, including gasoline, diesel, jet fuel, and lubricants. Additionally, this phase includes the establishment and operation of refineries, petrochemical plants, and retail outlets. Methane emissions in the downstream sector primarily originate from refineries, distribution systems, and retail outlets.

The O&G industry's upstream activities account for over 75% of total

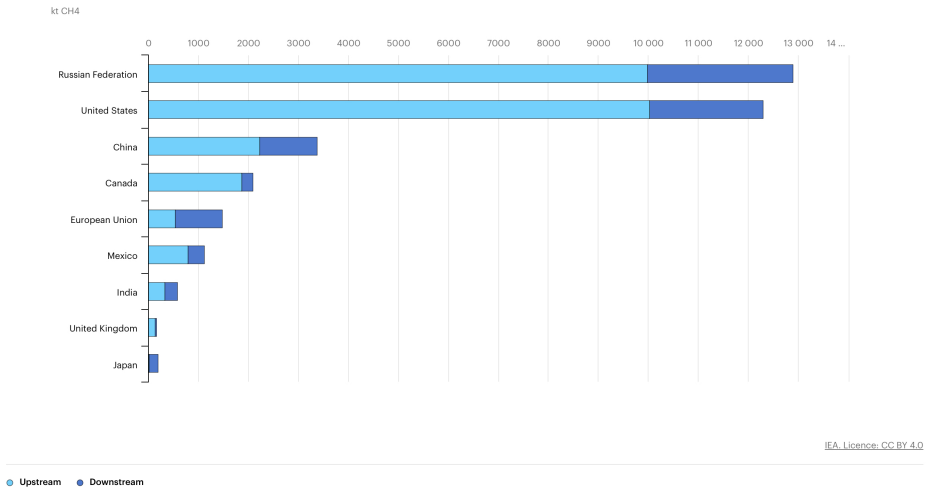


Figure 2.12: Oil and gas methane emissions in selected countries by oil and gas sector, 2020. *Image source: International Energy Agency (IEA).*

methane emissions, while the downstream segment contributes the remainder. As depicted in Figure 2.12, the United States recorded the highest methane emissions in the upstream sector in 2020, with approximately 10,000 kt emitted. In the downstream sector, Russia emitted nearly 3000 kt of methane, followed by almost 2500 kt in the United States.

The upstream and downstream sectors are structured around production, gathering, boosting, and processing sites. As shown in Figure 2.11, each site is equipped with specific infrastructure. The upstream sector consists of production sites, including wells, and gathering & boosting sites encompassing gathering pipelines, separators, compressors, pneumatic devices/pumps, storage vessels, heaters, and flares. The downstream sector comprises processing sites characterized by refineries and wells.

Each of these infrastructures contributes to varying levels of methane emissions. Studies [89, 156, 121, 48] have consistently identified storage tanks, flares, and compressors as prominent sources of methane emissions. However, the order of contribution can vary across different surveys. For instance, Figure 2.13 presents the findings of a recent aerial survey measuring methane emissions from various O&G infrastructures in British Columbia, Canada [49]. The study indicated that compressors accounted for 54% of the total methane emissions, followed by tanks at 18%, and unlit flares at

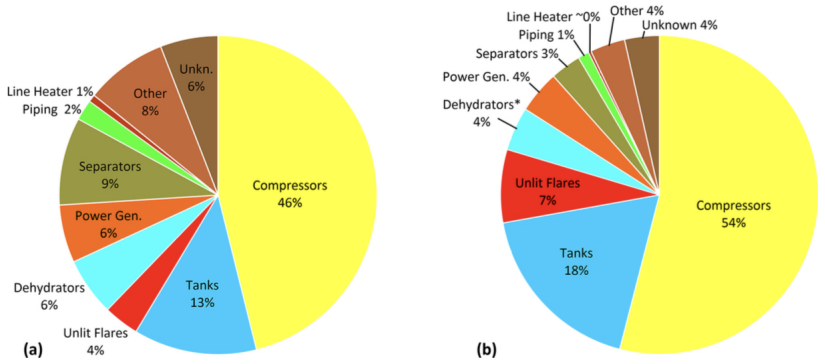


Figure 2.13: Relative contribution of source types to aerially detected methane. (a) Fraction of detected sources and (b) fraction of measured methane among quantified source types in the aerial survey[49].

7%.

Methane emissions from O&G production sites exhibit skewed distributions, where a small percentage of sites—often referred to as super-emitters—contribute to the majority of emissions. A recent study [15] found that these strong methane point sources contribute an average of 40% of total emissions across multiple basins in the United States, highlighting the disproportionate contribution of a limited number of emitters. Methane super emitters are facilities, equipment, or sources within the oil and gas industry that release significantly higher amounts of methane compared to the average emissions from similar sources. Identifying and addressing super emitters is essential for effective methane mitigation efforts, as they can significantly influence overall methane emissions from the oil and gas sector.

Super emitters can release large quantities of methane due to various factors, including equipment malfunctions, operational inefficiencies, or design flaws. Emissions from these super emitters, as well as from other sources, can be classified into three categories:

- **Fugitive** methane emissions result from unintended leakages, often due to faulty seals or leaking valves.
- **Vented** methane emissions are intentional releases, commonly for safety reasons, resulting from the design or operational requirements

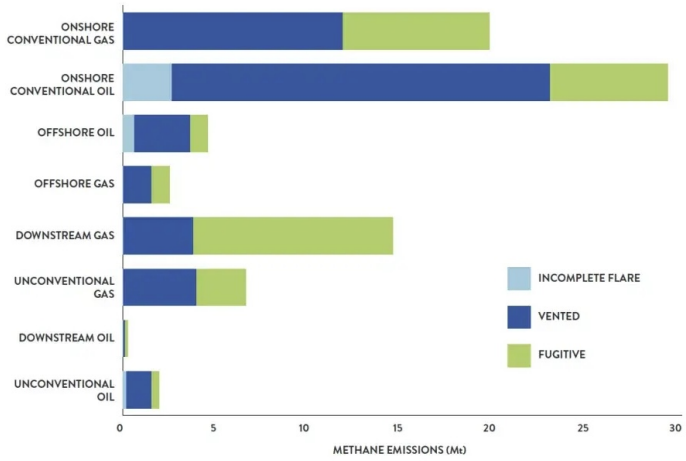


Figure 2.14: Annual oil and gas sector methane emissions by production type and reason in million tonnes (mt) in 2020. *Image source: International Energy Agency (IEA).*

of the facility or equipment (e.g., pneumatic controllers and flaring) or maintenance activities (e.g., venting a pipeline for inspection).

- **Incomplete flaring** methane emissions occur when natural gas, which cannot be economically used or recovered, is burned instead of being sold or vented. While the majority of the natural gas is converted into CO2 and water through combustion, a portion may not be completely burned, resulting in methane emissions being released into the atmosphere.

Figure 2.14 indicates that for oil and gas (on-shore and off-shore), vented methane emissions are predominant, signifying that these emissions arise from intentional actions. Regulatory measures tailored to discourage practices leading to methane emissions could mitigate vented emissions. For fugitive emissions and incomplete flaring, implementing new components and conducting repairs could offer mitigation solutions. According to the IEA, the global oil and gas sector emitted more methane in 2021 than Canada's annual consumption. If this methane were captured and sold at the current U.S. price of $4 per million British thermal units, it could generate approximately $17 billion. The IEA estimates that an $11 billion investment could eliminate around 75% of the world's methane leaks, as

well as a significant amount of gas wasted by flaring at the wellhead. Investing in repairs and infrastructure would not only reduce emissions and contribute to climate change mitigation but also generate profits for producers and provide additional natural gas.

2.4 Methane mitigation's actions

2.4.1 Current mitigation policies

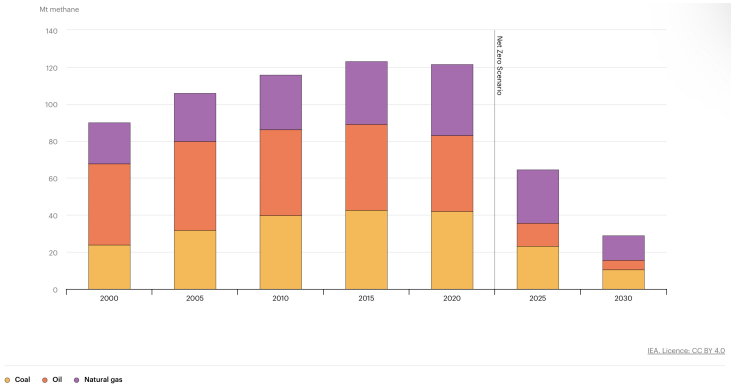


Figure 2.15: Methane emissions from fossil fuels, historical and in the Net Zero Scenario, 2020-2030. *Image source: International Energy Agency (IEA).*

Rapid reductions in methane emissions from fossil fuel activities are essential for achieving global climate targets, particularly when combined with substantial reductions in carbon dioxide emissions. Without targeted action on methane, the global average surface temperature could exceed 1.6°C by 2050, even with significant reductions in fossil fuel use. Under the Net Zero Emissions by 2050 (NZE) Scenario, total methane emissions from fossil fuel operations are projected to decrease by around 75% between 2020 and 2030 (Figure 2.15). Unlike agriculture and wetlands, the O&G industry has a long-standing history of reducing methane emissions due to safety concerns, and methane can be sold as natural gas, making emission reduction financially beneficial. The International Energy Agency (IEA) estimates that the industry can reduce its global emissions by 78%, and up to 39% [41] of these reductions can be achieved without financial compromise. These

reductions are crucial, as methane emission cuts from fossil fuel operations are projected to contribute half of the total reduction in methane emissions required by 2030 to limit warming to 1.5°C.

The oil and gas industry has instituted various policies and initiatives to mitigate methane emissions. These encompass:

- **Regulatory Measures**: Governments and regulatory entities across different countries have established regulations targeting methane emissions from oil and gas operations. These often entail provisions for monitoring, reporting, and mitigating methane emissions.
- **International Agreements**: International agreements and initiatives exist to address methane emissions from the oil and gas sector. For instance, the United Nations Framework Convention on Climate Change (UNFCCC) serves as a foundational treaty underpinning international climate negotiations.
- **Financial Incentives**: Certain governments and entities provide financial incentives to encourage methane emission reduction by companies. These incentives may comprise grants, tax credits, or other financial mechanisms.
- **Voluntary Initiatives**: Numerous oil and gas firms have voluntarily pledged to curtail methane emissions. These initiatives frequently entail establishing emission reduction targets, adopting best practices, and investing in innovative technologies. The Oil and Gas Methane Partnership (OGMP) exemplifies a voluntary initiative targeting methane emission reductions from oil and gas operations.
- **Technology Development**: Ongoing research and development efforts are directed towards devising new technologies and practices for methane emission reduction in oil and gas operations. This encompasses technologies for detecting and rectifying fugitive leaks, as well as for capturing and utilizing methane that would otherwise be vented or flared.
- **Public Reporting**: Some companies opt to publicly disclose their methane emissions, fostering increased transparency and accountability within the industry.

Mitigating methane emissions in the oil and gas industry lacks a one-size-fits-all solution. The efficacy of policy and regulatory frameworks is contingent upon jurisdiction-specific circumstances, encompassing political and

regulatory contexts, industry characteristics, emission source size and location, and policy objectives. Regulatory approaches exhibit varied advantages and disadvantages across jurisdictions, necessitating consideration of these factors in policy design.

National authorities possess multiple avenues for implementing policies and regulations aimed at reducing methane emissions from the oil and gas sector. These include:

- **Standards**: These encompass requirements for the adoption of particular technologies and operational practices, alongside quantifiable emission thresholds. Technical standards, often denoted as Best Available Technologies (BAT)[4], are prevalent, with emission limits frequently integrated with economic mechanisms such as emission fees or taxes. Requirements for regular leak detection and repair (LDAR) [142] [17] programs are also part of this category.
- **Economic instruments**: This category includes emission fees or taxes, emission trading systems, and offset credit schemes. Gas price and reforms in gas pricing can also fall under this classification.
- **Public-private partnerships and negotiated agreements**: These can manifest in diverse forms, ranging from loosely structured partnerships with voluntary objectives to formal agreements with compulsory regulations in the absence of meeting specified quantitative targets. Negotiated agreements might incorporate emission reduction objectives, an overseeing institution for managing and coordinating emission mitigation measures, and protocols for monitoring, reporting, and verifying adherence.

The United Nations Framework Convention on Climate Change (UNFCCC), established in 1992, aims to stabilize greenhouse gas concentrations in the atmosphere at levels that prevent detrimental anthropogenic interference with the climate system. The UNFCCC delineates several principles to direct the global response to climate change. A pivotal requirement of the UNFCCC mandates countries to formulate and regularly update national greenhouse gas (GHG) inventories. These inventories employ bottom-up methodologies (cf. Section 2.2.2), applying emission factors to source activity data. The compilation methodology for these inventories is stipulated by the IPCC Guidelines on National Greenhouse Gas Inventories. Such inventories offer a comprehensive account of all GHG emissions and removals within a country's boundaries, categorized by sector and activity. Utilized

to monitor progress towards national emission reduction objectives and to report emissions to the UNFCCC, these inventories also pinpoint emission sources and opportunities for mitigation, thereby informing the formulation of climate policies and measures.

In crafting methane mitigation regulations and strategies, policymakers heavily depend on these inventories. The precision in characterizing emission sources is crucial for the development of effective mitigation regulations and strategies.

2.4.2 Limitations and Challenges

Numerous studies [1, 11, 93, 100, 72, 104, 71, 146, 99] have highlighted systematic underestimation of methane emissions from the oil and gas (O&G) supply chain. This raises concerns regarding the reliance on bottom-up methodologies for estimating emission inventories, which subsequently inform regulatory guidelines for methane emissions reduction. A study by [97] indicated that IPCC Tier 1 emission factors underestimated methane emissions from the O&G sector in the United States. Such underestimation is likely to affect countries employing these emission factors in their inventories. The authors observed that updates to the IPCC Tier 1 emission factors are not anticipated in the foreseeable future. Additionally, [20] compared UNFCCC-reported emissions to a set of global inversions and found that some of the highest fossil methane-emitting countries report lower emissions to the UNFCCC than estimated by atmospheric inversions.

As outlined in Section 2.2.2, methane emissions exhibit considerable spatio-temporal variability influenced by factors like weather, operating conditions, and maintenance practices. Such variability may elude capture by bottom-up approaches, introducing inaccuracies in emission estimates. Furthermore, these approaches can be intricate and resource-intensive, necessitating comprehensive data collection and analysis, posing challenges for implementation, particularly in resource-constrained regions.

A substantial portion of the discrepancy in emission estimates stems from the omission of super-emitters from emission inventories [58]. These sources, characterized by elevated emission rates, account for an average of 40% of total methane emissions. However, pinpointing these super-emitters presents significant challenges due to their transient nature and resistance to standard leak detection and repair (LDAR) methods. Their intermittent

occurrence and varying locations further complicate identification.

The prevailing absence of reliable emissions data impedes governments' ability to implement targeted actions at the requisite scale and pace to realize the objectives of the Global Methane Pledge (GMP).

To address these challenges and align with GMP objectives, various initiatives and programs aim to furnish the most accurate and comprehensive methane emission data feasible. The Oil and Gas Methane Partnership (OGMP), a voluntary initiative spearheaded by the United Nations Environment Programme (UNEP) and the European Union, strives to curtail methane emissions from the oil and gas sector. A core OGMP objective is to prompt participating companies to transparently report their methane emissions and reduction endeavors, encompassing detailed data on emission sources, mitigation measures, and progress towards reduction targets.

OGMP offers five tiers of reporting methods, ranging from generic emission factors to on-site and infrastructure methane emissions measurements utilizing ground and aerial sensors (bottom-up & top-down).

From a bottom-up emissions perspective, on-site sensors have proliferated and enhanced in precision. Companies stand to benefit from these monitoring technologies by commercializing methane emissions data to interested entities, such as governments or other oil and gas companies. Furthermore, the deployment of ground-based emissions monitoring systems enables companies to devise effective methane mitigation strategies to circumvent fines levied under new methane regulations, like the Inflation Reduction Act (IRA) in the USA, which imposes charges per ton of CH_4 emitted exceeding 25,000 tons of CO2e. The evolution of the oil and gas industry towards Industry 4.0[70, 131, 75], underpinned by advancements in process automation technologies, entails augmented utilization of innovative sensors[74] like infrared cameras[24], LIDAR[76], and Laser Path, facilitating precise and quasi-continuous spatio-temporal representation of on-site methane emissions. Concurrently, various functional parameters associated with infrastructures, such as age, flow rates, maintenance, and operational efficiency of each equipment on site, will be monitored.

From a top-down emissions standpoint, the current satellite constellation comprises at least 15 satellites (for global & point source perspectives) dedicated to methane concentration monitoring. This satellite constellation is poised to expand through ongoing technological advancements offering enhanced spectral resolution, spatial resolution, and temporal cov-

erage (cf. Section 2.2.1). Furthermore, technological innovation in spaceborne methane measurement tools appears to prioritize long-term quasi-continuous monitoring, akin to advancements in carbon cycle analysis [45]. This regular methane emission measurement concept should facilitate continuous monitoring of emission levels from sites and infrastructures, enabling the establishment of their respective emission profiles. In the long run, the diverse range of measurement tools in operation, coupled with potential continuous measurement capabilities, is anticipated to generate substantial data volumes.

Confronted with the escalating volume of raw data from top-down and bottom-up sensors, the processing and analysis of this data pose significant challenges, particularly in real-time contexts where expeditious processing and analysis are imperative.

2.4.3 Solutions

To devise effective regulations for methane emissions reduction, it is essential to develop a robust and accurate inventory methodology that integrates multi-level methane emission profiles applicable to on-site equipment and national-level estimates.

The Methane Alert and Responses System (MARS) initiative, introduced by UNEP and the Environmental Defense Fund (EDF) at COP27 [57], aims to create a system capable of near real-time methane emissions detection from oil and gas operations. This system will leverage a combination of methane emissions data, including satellite data from the International Methane Emission Observatory (IMEO). The MARS initiative seeks to furnish actionable insights to operators and regulators to curtail methane emissions, facilitating enhanced monitoring and enforcement of methane regulations.

Such systems should be grounded on the deployment of comprehensive monitoring systems at both global and individual oil and gas infrastructure levels, enabling precise characterization of methane emissions across various levels (national, basins, operator, site, infrastructure). A recent study [15] found that, on average, 40% of emissions from US oil and gas fields emanate from point sources, underscoring the importance of characterizing methane emissions both as area and point sources [124]. Furthermore, identifying the characteristics of super-emitters and predicting the likelihood of

a site becoming a super-emitter, or forecasting the timing and location of super-emitter events, would significantly contribute to methane emissions reduction.

Characterizing methane emissions involves determining their source, volume, and location. Accumulated over time and across various levels, this data could facilitate access to spatio-temporal and time series data, serving as the foundation for in-depth analysis and spatio-temporal methane leak forecasting. This forecasting capability is pivotal for preventing methane leaks by providing a proactive window for intervention.

To enable real-time, global-scale operation of such a system, it must leverage automated methods for processing and analyzing the substantial volume of data implicated. The incorporation of artificial intelligence into this process is the key element capable of meeting this requirement.

Chapter 3

Automated Oil and gas infrastructures detection and recognition

This chapter introduces in Section 3.1 the general role of artificial intelligence and its state-of-the-art applications for diverse methane mitigation tasks. Section 3.2 focuses on the description of objection detection algorithms and their specific applications in oil and gas methane emissions mitigation domain. This chapter in Section 3.3 also presents a case study for the automatic detection and recognition of oil and gas infrastructures in the Permian basin (USA) based on the use of object detection algorithms and high resolution satellites images. Then, the results are presented in Section 3.4 including also the results algorithms pre-training effect and satellite images based adversarial attacks on our algorithms performances.

This chapter, is based on our 2 following publications : *"Oil and Gas Automatic Infrastructure Mapping: Leveraging High-Resolution Satellite Imagery through fine-tuning of object detection models"* [37] and *"Object detection models sensitivity & robustness to satellite-based adversarial attacks"* [36].

Contents

3.1 Artificial Intelligence for Methane Mitigation

3.1.1 Artificial Intelligence Introduction

The use and applications of artificial intelligence (AI) have experienced a massive increase over the last decade. AI has been fundamental for handling, in an automated way, heavy and repetitive time-consuming tasks which could require quasi-continuous attention and high precision. Machine learning (ML) and deep learning (DL) - which are AI subsets - allow for automated decision-making and forecasting, which are currently used in wide categories of applications such as medical diagnostics, speech recognition, recommendation system, and autonomous driving. ML can intervene in every domain, especially when dealing with large amounts of data and when analysis and decisions have to be done efficiently and in a short period of time.

AI enabled the development of supervised models capable of performing multiple prediction tasks from massive amounts of labeled data (Figure

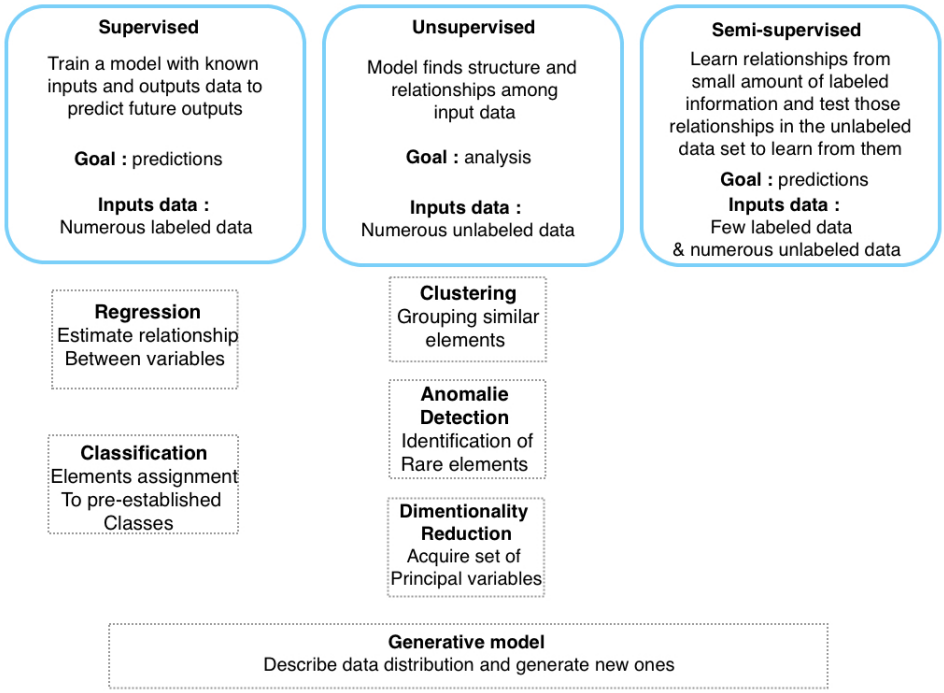


Figure 3.1: Machine Learning methods presentation.

3.1). However, data associated with real-life problems are not always labeled and sometimes requires long hours of manual annotation to be processed by supervised methods. In the case where a dataset is not labeled, the unsupervised or semi-supervised methods can be directly used without having to labeled data and just a small amount. However, unlabeled data can be manually or automatically annotated to achieve the necessary conditions for using a particular method.

As illustrated on the Figure 3.1, machine learning could regroups different methods for different tasks, for example :

1. **Regression** which is used to estimate relationship between variable and make forecasting;

2. **Classification** permits to automatically assign elements to classes;

3. **Clustering** is used to group elements without having pre-determined classes

For example, to perform an automatic methane plumes detection (e.g. classification) with very few plume images (semi-supervised), it is possible to extend the dataset (to be in a supervised learning setting) by automatically generating images similar to those in the original dataset. For this, generative models can be used to generate automatically new data. Generative models can therefore replace the heavy manual annotation work mentioned above and have the advantage of being applicable in supervised, semi-supervised, and unsupervised learning.

All these machine learning methods can be applied in various domains like Natural Language Processing (NLP) with, for example, text classification, natural language generation, automatic summarizing, etc. But also computer vision (images & videos) with classification, classification + localization object detection, instance & semantic segmentation for images and videos, 3D scenes generation, etc.

3.1.2 General State of the art

The use of AI in the context of methane emissions monitoring and O&G infrastructure has greatly increased in less than a decade. Indeed, the increasing number of aerial and ground sensors for methane emissions tracking produce large and diverse sets data. Facing this large amount of data, the use of automated methods to transform these data into interpretable informations is required. The Table 3.1 present various works based on AI by type of analysis :

1. Methane plume detection
2. Methane plume quantification
3. Methane emissions forecasting and O&G site/infrastructure detection and recognition

Methane plume detection. is part of the computer vision domain where images and videos could be automatically process in order to determine the presence or not of methane plume. Through the use of sentinel-2 images, [80] proposes a Convolutional Neural Network (CNN) method to automatically detect methane emissions and flaring activities. Still based on the use of sentinel-2 images, [129] through the CH4Net system, focus on the detection of methane super-emitter (large scale emissions) using a CNN U-Net model for binary segmentation pixel by pixel indicating the presence or not of a methane plume. [101] proposes a two-step machine learning approach also using a CNN to detect plume-like structures in the methane data and subsequently apply a support vector classifier to distinguish the emission plumes from retrieval artifacts based on TROPOMI images (Sentinel-5P). The GasNet method [135] demonstrates that methane plume could also be detected from optical gas videos using CNN models applied on different background subtraction methods.

Methane plume quantification. relies on chemical transport models (cf. Section 2.2.2) which requires the use of local measurements such as background wind speeds. The MethaNet methods[50], based on CNN, proposes to predict (regression) methane point-source emission directly from high-resolution 2-D plume images without relying on other local measurements from Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG). With the same objective, [87] compares different CNN architectures such as VGG-19, ResNet-50, Inception-v3, DenseNet-121, Swin-T, and EfficientNet-V2L for methane plume quantification based on Sentinel-2 images. Alternatives approaches based on classification such as [133] and VideoGasNet [136], make it possible to classify methane emissions by sizes from optical gas video with the use of CNN model's Long Short Term Memory network (FC-LSTM) and the convolutional LSTM (ConvLSTM).

The literature also features studies combining both detection and quantification of methane plume, like [109] where the authors compare the use of neural network and reduced support vector machine (RSVM) models based on images from unmanned aerial vehicles (UAVs) and atmospheric parameters. [8] presents a two-step algorithm called U-Plume for automated detection and quantification of point sources from satellite imagery from GHGSat-C1. The first step delivers plume detection and delineation (masking) with a machine learning U-Net architecture for image segmentation. The second step quantifies the point source rate from the masked plume using wind speed information and either a convolution neural network (CNN) or a physics-based Integrated Mass Enhancement (IME) method. [53] also

proposes a CNN based method for detection and quantification from gas pipelines in the subsea with the use of simulated flow and its parameters.

Methane emissions forecasting could be realized for different tasks such as spatio-temporal forecasting, probabilistic classification or simple forecasting. [134] proposes to predict high-emitting sites that can be prioritized for follow-up repair based on several variables contribute to the formation of leaks such as infrastructure age, production, weather conditions, and maintenance practices and methane emissions quantities. This study compares various regression methods such as Logistic Regression, Decision Trees, Random Forests and AdaBoost. [52] proposes to estimate the emission fluxes of methane from open-pit mines based on the uses of multi-layer perceptron (MLP) artificial neural network, the gradient boosting (GBR), XGBOOST (XGB), and support vector machines (SVM) with simulated emissions flux and associate meteorological parameters. [116] compare the use of SVM and back propagation neural networks for pipeline methane leak forecasting based on the simulation of low-pressure gas pipeline system, which reproduces the gas leakage scenes and its associates parameters like pressure wave. The literature also contains hybrid approaches combining physic/chemistry models with neural network models. [106] proposes a hybrid deep learning model by integrating variation inference and physical constraint with a deep learning backbone to forecast spatio-temporal concentration evolution of natural gas release at plume area.

Others. tasks based on artificial intelligence could also contribute to the reduction of methane emissions. For example, [59] proposes a solution to detect, locate, and estimate the size of O&G pipelines metal-loss defects with the use of dimension reduction through Principal Component Analysis (PCA) and neural network. The placement of ground sensors also plays an important role for methane emissions monitoring, [138] proposes to optimize sensor placement to maximize the detection of possible methane leaks. To this end, the authors present a machine learning approach based on the use of the clustering algorithm DBSCAN which leverages various data sources including oil and gas facilities data, historical methane leak rate distribution and meteorological data. An other study [132] proposes a techno-economic analysis based on the comparison of classical Leak Detection and Repair (LDAR) and machine learning based LDAR for automated leak detection with the use of neural network and optical gas imaging. They highlights the benefits of automated LDAR including reduction in labor cost, prioritization of large leaks, and maximization of mitigation efficiency

and minimization of the net cost of mitigation.

All these studies, using artificial intelligence methods, makes a significant contribution to improving the monitoring of methane emissions and its mitigation. However, in order to achieve optimum mitigation, the behaviour of methane-emitting sources must be precisely characterised so that a maximum of unintentional emissions can be avoided. To reach this goal, fine-grained automated monitoring of emitting sources is required. This process begins with the automatic detection and recognition of sources (Section 3.2), then the spatial matching of the sources at the origin of the detected emissions (Chapter 4).

3.2 O&G facilities detection and recognition

3.2.1 Object detection State of the art

Object recognition algorithms, a subset of computer vision techniques, facilitate the automatic identification and location of multiple instances of a given class of objects in images or videos. These algorithms use both neural network-based methods or non-neural techniques. The non-neural strategies typically integrate SIFT [68] or HOG [16] (for feature extraction) with a classification algorithm such as Support Vector Machines (SVM). Despite their usefulness, recent studies suggest that neural-based object recognition methods generally outperform their non-neural counterparts [137]. Neural approaches to object recognition can be divided into three categories depending on the degree of supervision in their learning process: supervised, semi-supervised, and self-supervised models [40]. The works conducted in this thesis only focus on the use of supervised algorithms. Supervised object recognition algorithms require an annotated image database for effective training. In this context, annotating an image involves identifying objects of interest by enclosing them in a bounding box and labeling them appropriately. During supervised learning approach, an object recognition algorithm learns to locate and subsequently recognize the targeted objects. The structural organization of object detection algorithms is usually defined by the following main components:

- **Backbone**: This refers to a deep learning architecture, usually a convolutional neural network (CNN), that is tasked with the essential

function of feature extraction. Through this process, the backbone identifies and abstracts the salient features from the input data;

- **Neck**: Serving as an intermediary between the backbone and the head the neck performs a fusion of the features extracted from the different layers of the backbone model. This synthesized information forms the basis for the subsequent predictions performed by the head;
- **Head**: The head forms the final component of the object recognition model and is responsible for predicting the classes and bounding box regions. These predictions form the final output of the object recognition model. In particular, the head can produce a number of outputs, typically configured to detect objects of different sizes in an image.

Independently from learning type, this process can be executed over three primaries architectural frameworks [145]:

- **Two-stage detectors:** is based on two main models, firstly Region Proposal Network (RPN) which are fully convolutional networks used to extract regions of objects, and secondly an extra model is used to classify and further refine the localization of each region proposal. RCNN [31] architecture is based on a selective search algorithm to propose regions of interest and then applies a CNN to each region to classify it as an object or background. As this method was considered as particularly slow, the Fast-RCNN [30] authors proposed, an optimized approach to RCNN by sharing computation across all regions proposed in an image. Finally, FASTER-RCNN [92], based on the architecture of Fast-RCNN, replaces the selective search algorithm with a RPN, which is trained to directly predict regions of interest. This latest version reduces computation time and improves the detection accuracy;
- **One stage detectors:** Contrary to two stage detectors, one stage detectors don't need to integrate RPN to generate a region proposal, it can directly obtain the classification accuracy of the object and its coordinate position. These algorithms have the advantage of being faster than two-step algorithms. In this category, there are YOLO [91] and its different versions [113], SSD [69] and RetinaNet [66]. Review studies compares the latter 3 methods. For example [112] for the pill identification task, showing that YOLO v3 offers the best performance in terms of execution time but the lowest accuracy. Another study

[23] focuses on the comparison of SSD, RetinaNet, YOLO v4 and FASTER-RCNN for Tethered Balloon detection. It was shown that YOLO v4 achieved the best trade-off between speed and accuracy with a precision of 90.3%. [46] also concludes that YOLO has better accuracy (increasing with version) via a broad comparison of RCNN and YOLO models and their variants;

- **Others:** There are also object detection methods based on approaches other than the one-two stage approaches detailed above. For example, DETR[10] is a transformer based detector with a 3 parts architecture constitute of a CNN, encoder-decoder transformer and a feed-forward network (FFN).

Backbone Pre-training. The majority of object recognition models, including but not limited to YOLO, FASTER-RCNN, and DETR, provide an option for a pre-trained version of the backbone. This pre-training generally helps to improve recognition performance. The pre-training of these algorithms is done using extensive databases of thousands of image categories, ranging from everyday objects such as airplanes, dogs, chairs, etc. Prominent among these databases are ImageNet [19], which contains 200 classes and about half a million annotated objects, and COCO [67], which contains 80 classes and nearly 1 million annotated objects. In addition, the Pascal database VOC [26] includes about 20 classes with about 63,000 annotated objects. Most recent object recognition algorithms are pre-trained on the COCO dataset. A main advantage of pre-training backbones is the significant reduction in the custom dataset training phase. Pre-trained backbones that have already learned to recognize general features and patterns from large databases can transfer this knowledge to the object recognition task at hand. This not only minimizes training time, but also enables the use of smaller datasets. Pre-built models also play a critical role in mitigating the problem of over-fitting, which occurs when the model over-learns from the training data, compromising its ability to generalize to new data.

3.2.2 Object detection O&G applications

Remote sensing object detection can be applied to a variety of problems, various studies [7, 139, 65, 51] summarizes object recognition algorithms applied to various remote sensing topics. For example, [63] summarizes the performance of FASTER-RCNN, SSD, and YOLO v3 algorithms for agri-

cultural GHGs detection based on high-resolution satellite imagery. [38] proposes automatic detection of earthquake-induced ground failure effects by using FASTER-RCNN. Others [18, 64, 111]focus on comparing one and two-stage object detection algorithms on satellite and aerial images. [154] uses DETR for object detection with enhanced multi-spectral feature extraction. In particular, object detection algorithms are also used for problems in the oil and gas sector. For example, in the works [144] [143] YOLO v4 is used to detect oil spills with Sentinel-1 SAR images. Some studies are also looking at oil and gas infrastructure detection:

- **Oil Tanks**: [148] proposes a recognition algorithm that harnesses deep environmental features, using the convolutional neural network (CNN) model and SVM classifier for oil tank recognition. Another study employs FASTER-RCNN for the same objective;
- **Oil Wells**: [107] introduces an enhanced version of YOLO v4 for detection using high-resolution images, similar to [108], where the authors utilize FASTER-RCNN algorithm. [140] presents a database, dubbed Northeast Petroleum University-Oil Well Object Detection Version 1.0 (NEPU-OWOD V1.0), which includes the geographical locations of oil wells. This database was constructed via the application and comparison of nine object detection algorithms;
- **Pipelines**: In the context of pipelines, [28] uses a deep learning approach for object detection in underwater pipeline images, employing various YOLO configurations;
- **Oil & Gas Sites**: On a broader scale encompassing entire infrastructures, [149] employs high-resolution satellite images and YOLO v2 for automatic recognition of oil industry facilities, with a particular emphasis on well-sites.

In the field of object detection, a significant portion of existing methods are dedicated to the identification of specific infrastructures. While this focused approach proves beneficial in studies examining a single infrastructure, it may not be entirely sufficient when examining methane emissions in the oil and gas (O&G) sector. This is because such emissions can come from a variety of infrastructures. Recognising this multi-faceted challenge, our work broadens its scope to include three types of infrastructure that are essential to the O&G sector: Wells, Tanks, and Compressors (cf. Subsection 2.3.2).

3.3 Permian basin infrastructures detection

3.3.1 Introduction

To effectively reduce these emissions in the O&G sector, a comprehensive understanding of the emissions profiles of individual operators, specific sites, and associated infrastructure is needed. This knowledge would inform the formulation and refinement of regulatory measures and potential penalties to ensure they are appropriately tailored and thus optimally effective.

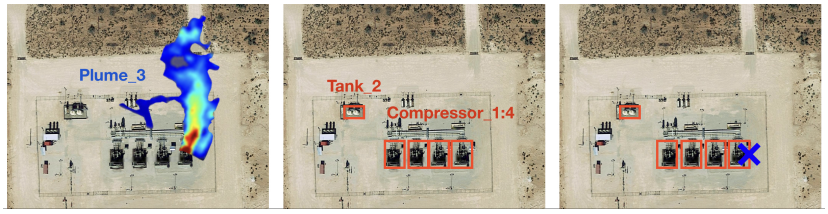


Figure 3.2: Example of a detected methane plume associated with infrastructure at its source. In 3 automatic steps: detection of methane plume, detection of infrastructure, association of each plume and infrastructure. *Source: @Google Earth.*

In order to characterise the emissions from each infrastructure, each methane plume detected must be automatically associated with the infrastructure from which it originated as illustrated on the Figure 3.2. If we consider a case where for each detected methane plume a satellite image is extracted representing site and infrastructures, these infrastructures have then to be automatically detected and in a second place one of these last ones have to be spatially associated with methane plume.

This Permian basin study case focuses on the automated detection of oil and gas infrastructures, a topic that has not been explored in depth in the existing literature. Sites in the oil and gas industry that contain wells, storage tanks, or compressor infrastructures (Figure 3.4) are considered significant contributors to fugitive emissions and therefore form the targets we seek to automatically identify. Existing approaches to oil and gas infrastructure detection typically do not allow for the simultaneous detection of multiple infrastructures.

Our work focus on the detection and recognition of O&G infrastructures in

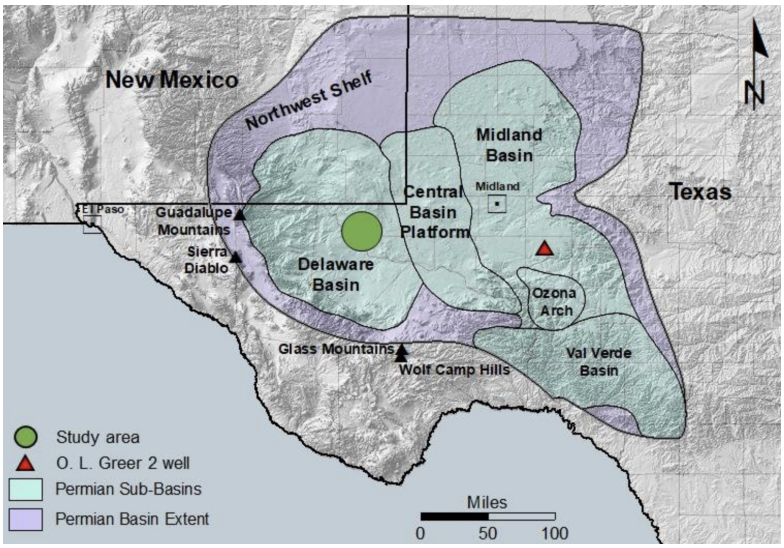


Figure 3.3: The Permian Basin extends over two states, Texas and New Mexico, which is divided into multiple sub-basins including Delaware and Midland basin. *Source: [90].*

the Permian basin in USA (Figure 3.3). The Permian basin is the America's largest and most productive oil basin, it extends over 168,000 km2 in the Texas and New-Mexico states.

With the goal of enabling the automatic detection of compressors, tanks, and well infrastructures simultaneously, we have used supervised object detection methods, specifically using and comparing (cf. Subsection 3.4) the YOLO, FASTER-RCNN, and DETR object detection algorithms (cf. Subsection 3.3.3). These algorithms, initially pre-trained on the COCO database, are fine-tuned using the Oil and Gas (OG) database (cf. Subsection 3.3.2). The OG database, which was developed specifically for this study, contains satellite view with high spatial resolution (less than 1 meter). The images in the OG database are randomly extracted from the Permian Basin. We also focus on the impacts of pre-training on algorithm's results by comparing our 3 algorithms results with and without pre-trained networks (cf. Subsection 3.4.1). Finally, we also present and test the impacts of potential satellite images noises on the results of our object detection algorithms, permitting to compare their robustness to different satellite image noises (cf. Subsection 3.4.2).

3.3.2 O&G benchmark dataset

Algorithms employed in supervised object recognition require a learning phase involving substantial interaction with a large repository of images. These images must be labelled with the target object in order to enable practical training. In the specific context of identifying wells, tanks, and compressors, this database must contains a variety of aerial photographs in which each of these objects or infrastructures is unambiguously identifiable. The procurement of such specialized labelled images, due to the lack of public availability, demanded the development of a dedicated database specifically designed for this purpose.

Figure 3.4: Example of images and annotated objects from OG database : tank (red), compressor (purple) and well (blue)*source : @Google Earth.*

In this study, we chose to extract high-resolution satellite images from the Permian Basin region (over the states of New Mexico and Texas in the USA), which is the largest O&G basin in the world. 930 Google Earth images of O&G sites with their infrastructures were extracted, with resolutions ranging from 15cm to 1m. Each of these images was then manually annotated by drawing bounding boxes around each well, compressor or tank present, as shown in Figure 3.4. Each of these boxes is associated with 1 of our 3 objects (label). In total, out of the 930 images, 1951 objects were annotated: Compressor 706 objects, Well 630 objects and Tank 615 objects

All the images are in a 640x640 size format, each featuring between one and multiple instances of key infrastructure such as wells, tanks, and compressors. By site, and therefore per image, compressors appear in groups, generally between 2 and 5 as shown in Figure 3.3.2. Concerning tanks, in the database we have considered tank units and not individual tanks, which explains why only one unit is generally present on an image (Figure 3.3.2). And finally, wells are in the vast majority of cases present alone on a site (Figure 3.3.2).

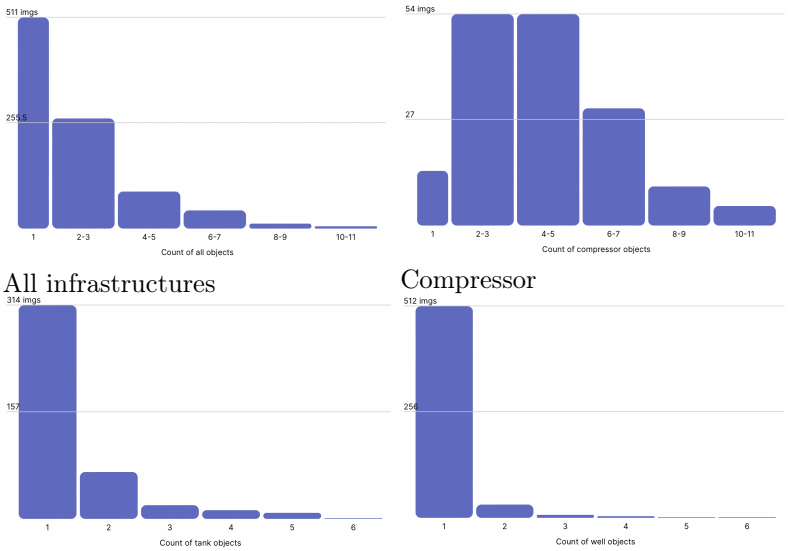


Figure 3.5: Number of images in function of number of infrastructures in the OG database.

Another aspect worth mentioning is the special consideration given to wells in our database. Given the limited resolution of satellite imagery, it is often difficult to discern the structural details of wells. Therefore, the recognizable shadows of wells that are present even at lower resolutions are included in the bounding boxes (as shown in the right column of Figure 3.4). This database is hosted on the open-source Roboflow platform and can be accessed via the following link: https://universe.roboflow.com/thesis-ffaad/og-otgc5/dataset/6. Following the requirements of a rigorous study design, we have divided our dataset into different subsets for training, validation, and testing. Of the total images, 80% (744 images) are used for training, 13% (or 120 images) for validation, and the remaining 7% (66 images) for testing.

3.3.3 Object detection algorithms

The 3 algorithms FASTER-RCNN, YOLO and DETR selected for this study each have different architectures but also different pre-trained models (backbones). This Section present their details and also the models

parameters selection for their fine-tuning :

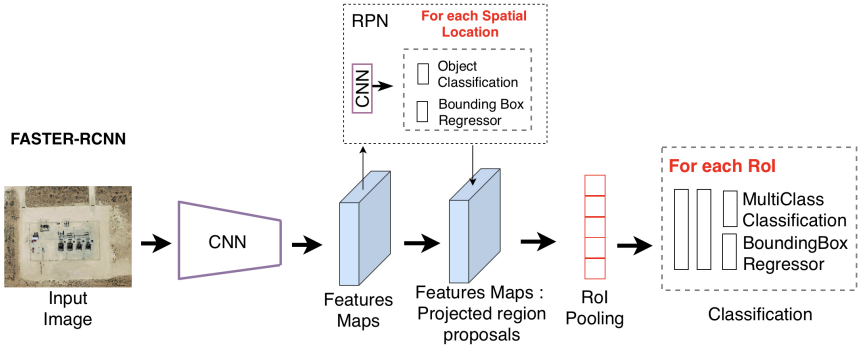


Figure 3.6: FASTER-RCNN architecture.

FASTER-RCNN [92] process in 2 main steps, first it uses use a Region Proposal Network (RPN) to generate regions of interests and secondly it send the region proposals down the pipeline for object classification and bounding-box regression.FASTER-RCNN architecture is based on 3 principal components : the backbone (CNN type varies according to chosen model), the RPN , and the ROI heads (classification and regression). FASTER-RCNN provide 3 backbones architectures pre-trained on COCO 2017 base (train2017 and val2017) :

- **Feature Pyramid Network (FPN)**: Use a ResNet+FPN backbone with standard conv and FC heads for mask and box prediction;
- **C4**: Use a ResNet conv4 backbone with conv5 head which correspond to the original baseline in the FASTER-RCNN paper;
- **Dilated-C5 (DC5)**: Use a ResNet conv5 backbone with dilations in conv5, and standard conv and FC heads for mask and box prediction, respectively.

We have fine-tuned 2 FPN model with ResNet50 and ResNet101, but also a DC5 model based on ResNet101. Epochs were fixed to 100, batches to 64 and learning rate to 0,001.

You Only Look Once (YOLO) [91] v8 is one of the most recent versions which outperforms previous versions in term of precision as illustrated in Figure 3.8.

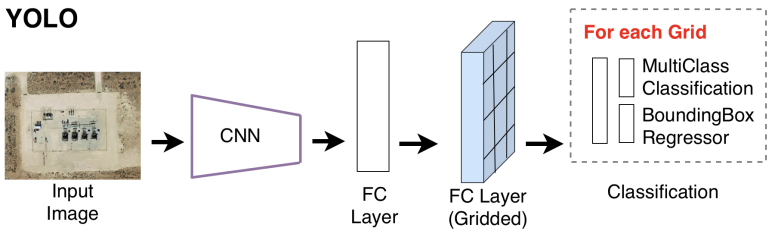


Figure 3.7: YOLO architecture: Convolutional Neural Network (CNN), Fully Connected (FC) layer, Girdded FC layer.

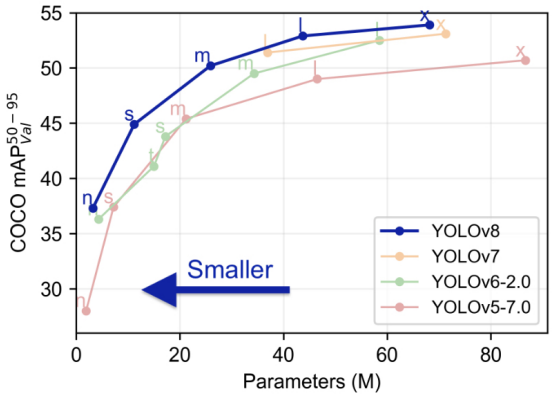


Figure 3.8: YOLO Mean Average Precision (mAP) for COCO object detection by versions and models. *Source :* *https://github.com/ultralytics/ultralytics*.

YOLO v8 utilizes a modified version of CSPDarknet53 as its backbone, featuring 53 convolutional layers and cross-stage partial connections to enhance information flow. The head of YOLO v8 comprises convolutional layers followed by fully connected layers, responsible for predicting bounding boxes, objectness scores, and class probabilities. Incorporating a self-attention mechanism, YOLO v8 dynamically adjusts feature importance, enabling multi-scaled object detection through a feature pyramid network, detecting objects of various sizes within an image. Its backbone is then fine-tuned on the COCO dataset to learn to detect objects in 80 different categories. YOLO v8 has a declination of 5 pre-trained models (n,s,m,l,x) trained on COCO 2017 dataset. These models vary according to the number of parameters they hold directly influencing the level of precision, thus,

the more parameters a model has, the better its accuracy (cf. Figure 3.8). The 3 last pre-trained models (m,l,x) with the highest number of parameters were chosen and fine tuned with the OG database (image size 640x640), with 100 epochs, 16 batches, learning rate 0.001.

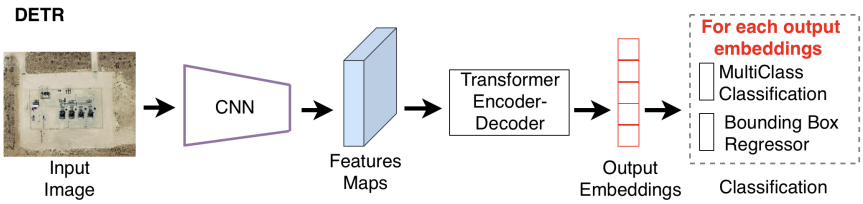


Figure 3.9: DETR architecture.

There are also object detection methods based on approaches other than the one-two stage approaches detailed above. For example, DETR[10] is a transformer based detector with a 3 parts architecture constitute of a CNN, encoder-decoder transformer and a feed-forward network (FFN).

Unlike one-stage and two-stage detectors, DETR [10] is designed as a direct set prediction problem encompassing a unified architecture. DETR employs a backbone (with varying architecture contingent on the selected model), a transformer encoder-decoder architecture, and a bipartite matching between predicted and ground-truth objects. By uniting the backbone and transformer, DETR successfully simplifies the architecture by eliminating specific components to one and two-stage approaches such as anchor generation and non-maximum suppression (NMS). The following pre-trained backbone models are available, all of which have been pre-trained on the COCO 2017 database:

- **R50**: Incorporates a backbone that is based on an ImageNet pre-trained ResNet-50 model.
- **R101**: Deploys a backbone grounded in an ImageNet pre-trained ResNet-101 model.
- **R50-DC5**: Increases the feature resolution by employing dilation in the final stage of the backbone. The backbone in this model is based on ResNet-50.
- **R101-DC5**: Implements a similar process to R50-DC5 but relies on a backbone built on ResNet-101.

Pre-trained model R50, R101 and R101-DC5 were selected for test with epochs fixed to 100, batches to 2 and learning rate to 0,001.

Model evaluation. Average precision (AP) is a widely used metric for evaluating the efficiency of object recognition tasks. The AP combines the precision and recognition curves into a single scalar quantity. The AP value ranges from 0 to 1 and tends toward 1 when both precision and recall are high, while it tends toward zero when either metric is low over a spectrum of confidence thresholds. AP is computed by calculating the difference between the current and subsequent recalls and multiplying that difference by the current precision:

$$AP = \sum_{k=0}^{k=n-1} [\text{Recalls}(k) - \text{Recalls}(k+1)] \times \text{Precisions}(k)$$

Where k is the number of object and n is the number of threshold. In addition, the mean average precision (mAP) is often used. It represents the average of AP calculated over all classes:

$$mAP = \frac{1}{n} \sum_{k=n}^{k=1} AP_k$$

where AP_k is the AP of the class k and n the number of classes.

Environment. For each algorithms (YOLO v8, FASTER-RCNN, and DETR), 3 models with different parameters and architectures were selected and compared. For each models, the output corresponds to the AP by class (Compressor, tank and well) and the mAP for the general model. The experiments were conducted with the use of a GPU NVIDIA GeForce RTX 3090 with 24 GO of memory. The experiments required the use of 3 distinct environments for the 3 algorithms with the following packages. YOLO v8: ultralytics (Python 3.8 environment with PyTorch 1.8). FASTER-RCNN: Detectron2 with torch 1.5 and torchvision 0.6 and DETR with PyTorch 1.7 and torchvision 0.7.

3.4 Results

To facilitate visual inspection of the previous results, the pre-trained model with the highest average precision (mAP) was selected for each algorithm and tested against the Oil and Gas (OG) database test data. Figure 3.10 illustrates the recognition performance of each algorithm model when applied to four different images from the test data.

Figure 3.10: Visual object detection results from pre-trained YOLO v8, FASTER-RCNN and DETR on 4 test images from OG database (images sources @Google Earth).

Case A: This scenario showcases a view of three wells that are closely spaced. YOLO v8 is able to detect and correctly discriminate each well. In contrast, DETR detects the presence of wells but combines the first two into a single object. FASTER-RCNN, which has significantly low average precision (AP) for the well class (as shown in Table 3.2), is not able to detect a well in this particular scenario.

Case B: This case represents an unusual circumstance where the appearance of compressors is underrepresented in the Permian Basin and consequently in the learning database OG (the most common representation is

shown in Case B). YOLO v8 can only detect one of the five compressors and one of the three tank units. DETR shows a slight improvement and detects three of the five compressors and all three tank units. Interestingly, FASTER-RCNN shows superior performance accurately recognizing all infrastructures without error.

Case C: This scenario presents a view of a typical compressor type found in the Permian Basin. In this specific instance, all algorithms correctly identify the six compressors and a single tank unit.

Case D: This case showcases a view of two sites each with a well. Unlike Case A, the image resolution in this case is lower and the wells are more widely spaced. YOLO v8 and DETR successfully recognize the two wells, while FASTER-RCNN fails to recognize either.

As highlighted in Case B, the visual representation of compressors in the Permian Basin is variable. The OG database contains a few cases where compressors are protected by a roof (as shown in the Case B images). To evaluate the detection capabilities of the algorithms in these particular circumstances, tests were extended to three additional images from the test database showing covered compressors (see Figure 3.11 results).

Case E: This scenario involves two unusual compressors along with a tank unit. FASTER-RCNN manages to identify all infrastructures, but it also mistakenly recognizes an additional compressor. DETR delivers an intriguing result by identifying a compressor through only a small segment protruding from the roof, as well as the tank unit. As for YOLO v8, it only manages to recognize the tank unit.

Case F: This scene provides a view of two unusual compressors and a tank unit. Both YOLO v8 and DETR fail to recognize the compressors, with DETR only acknowledging the tank unit. Yet again, FASTER-RCNN successfully identifies all the infrastructures as expected.

Case G: This scene presents a view of two unusual compressors exclusively. YOLO v8 is unable to detect either of them, while DETR correctly identifies one of the two compressors. It also detects an additional one, which does not correspond to a compressor but rather a small piece of infrastructure. Finally, FASTER-RCNN is also able to correctly recognize both compressors in this case.

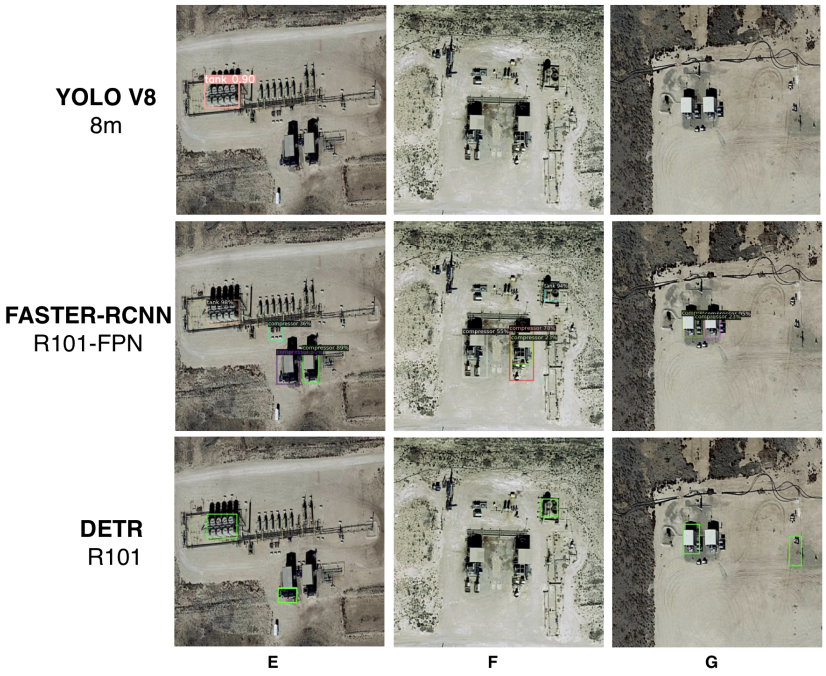


Figure 3.11: Comparison of pre-trained YOLO v8, FASTER-RCNN and DETR on a special (less representative) compressor architecture in the Permian Basin (images sources @Google Earth).

3.4.1 Pre-training effect

In general, pre-trained models offer numerous advantages over non-pre-trained models, including the need for less data (fine-tuning) and an improvement in accuracy. The models selected for this study were trained using the COCO dataset, which consists of annotated everyday objects. However, the COCO dataset does not contain any objects from industry that could resemble the objects in the OG database. This discrepancy raises the question of the extent to which pre-trained models, originally trained on objects that are significantly different from our target objects, can still outperform the predictive accuracy of non-pre-trained models. To investigate this, the algorithms and models discussed previously were run without the weights from the pre-training phase.

According to the results delineated in Table 3.3, it is observed that the

mean Average Precision (mAP) for the non-pre-trained YOLO v8 models is marginally lower than that for the pretrained models. Indeed, the average mAP of the YOLO v8 models is 89.4%, while that of the non-pre-trained models is 88.6%, indicating an overall decrease in mAP of 0.8%. While the difference may seem negligible, this result substantiates the assertion that pre-training YOLO v8 contributes to enhanced (slightly, in our results) performance.

In compliance with the results delineated in Table 3.3, we observe that the mean average precision (mAP) for the non-pre-trained YOLO v8 models is marginally lower than that of the pre-trained models. In fact, the average mAP of the YOLO v8 models is 89.4%, while that of the non-pre-trained models is 88.6%, which represents an overall decrease in mAP of 0.8%. Even though the difference seems negligible, this result supports the claim that YOLO v8 pre-training contributes to improved performance.

In terms of average precision (AP) by class, the results mirror those of the pre-trained YOLO v8 models; AP remains higher for the compressor class and lower for the well class. An interesting observation is the comparative analysis between the non-pre-trained YOLO v8 model and the pre-trained FASTER-RCNN and DETR models. The non-pre-trained YOLO v8 model outperforms all pre-trained FASTER-RCNN and DETR models in terms of mAP. This remarkable result demonstrates the superior efficacy of YOLO v8.

As regards of FASTER-RCNN, the pre-trained R50-FPN model shows significantly low AP and mAP. For the other models of FASTER-RCNN and all DETR models, convergence proved difficult even after increasing the number of iterations 20-fold and decreasing the learning rate by a factor of 1000. Non-pretrained models are notorious for their difficulty in achieving convergence, especially when dealing with smaller databases. The OG database is comparatively small, which may explain the observed lack of convergence, especially when compared to the larger COCO database.

3.4.2 Object detection robustness to satellite based adversarial attacks

While the 3 models fine-tuned with satellite imagery from the OG database — in particular YOLO v8 and RT-DETR — show promising accuracy, it is crucial to assess their robustness for practical remote sensing.

In supervised learning, a database of images labeled with objects is created so that these algorithms can learn to recognize and to retrieve specific objects. However, a notable challenge with this approach, especially in the context of remote sensing, is the limited representational diversity of the objects in the training database. Indeed, satellite images are well-known to exhibit a wide range of variations due to different satellite sensor types (e.g. resolution, noise) and environmental factors related to the position of the satellite and the sampled area (e.g. tangential distortion, brightness and saturation) [88]. These variations can significantly affect the algorithm's ability to generalize across different representations of the same object [86, 155].

We evaluated the robustness of YOLO, FASTER-RCNN, and RT-DETR against five specific types of variations commonly found in satellite imagery. To achieve this, we used the concept of adversarial attack during the inference phase as described in previous studies [147, 62, 151, 110]. In this method, the accuracy of both the pre-trained and fine-tuned algorithms is evaluated when they are presented with modified images (counterexamples) containing different types of induced perturbations. For each of the five types of variations, we systematically generated step-wise negative examples and then tested the performance of the algorithms for each sub-variation.

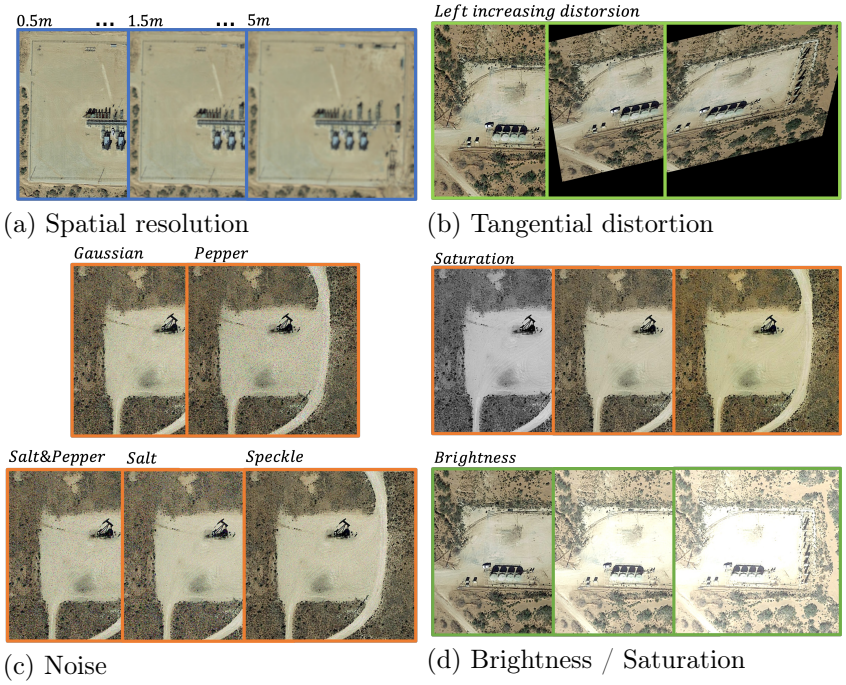
(a) Tank (b) Compressor (c) Well

Figure 3.12: Base test images of O&G infrastructures. *Source : @Google earth.*

Satellite images are subject to a number of variations that affect their visual representation. Technological differences between satellite sensors can lead to variations in spatial resolution and noise levels. Also, systematic errors can occur due to the Earth's rotation, geometric distortions caused by topography shifts, variations in satellite altitude and attitude, and instrument anomalies. Furthermore, environmental factors such as the nature of

the terrain (e.g. deserts) can influence image attributes such as luminosity and saturation.

To assess the robustness of our algorithms against variations in the satellite images, we implemented adversarial attacks [147, 62, 151, 110]. In this method, the precision of the 3 pre-trained and fine-tuned algorithms is evaluated in response to various perturbations generated as negative examples on the input images. Such a process is crucial to identify weaknesses in the models and suggest areas for refinement. For example, a notable drop in precision or a complete failure to recognise an object when confronted with certain adversarial examples could indicate a lack of resilience to that specific variation. We generated adversarial examples for each type of "base" image that mimic variations in satellite imagery, such as changes in resolution, tangential distortion, noise, brightness, and saturation.



(a) Spatial resolution (b) Tangential distortion

(c) Noise (d) Brightness / Saturation

Figure 3.13: Example of different degrees of satellite noises generated for O&G infrastructures images.

Spatial resolution variations. Different satellites, equipped with different imaging systems and specifications, have different spatial resolutions.

These differences result from factors such as sensor types, technological advances and the intended functions of the satellites. To mimic these resolution differences, as shown in Figure 3.4.2 (a), we applied average pooling with a square window (kernel size of 5 and stride of 3). In this way, ten images with resolutions ranging from $0.5m$ to $5m$ were generated, which are shown in Figure 3.12. This range mirrors the resolution spectrum of actual satellite images and thus allows a thorough evaluation of the performance of the object detection algorithms at different spatial resolutions.

Tangential Distortion. Tangential scale distortions, common in satellite imagery, result from the compression of image features, especially those further from the nadir point. Factors contributing to this distortion include sensor optics, the scanning system's motion, Earth's curvature and rotation, and terrain relief variations. To emulate varying levels of this distortion as depicted in Figure 3.4.2 (b), we employed a left affine perspective transformation using the OPENCV library. This technique involved a transformation matrix that maps 3 points from the original image to their new positions in the distorted image. We progressively adjusted the value of the left element in the matrix's third line from 50 down to 0, in steps of 5. This method enabled a controlled increase in leftward distortion across the images, providing a framework to evaluate the algorithms' effectiveness under different degrees of tangential scale distortions.

Noise. As shown in Figure 3.4.2 (c), the different on-board sensors of each satellite can cause a range of noise types. These include additive (Gaussian), multiplicative (speckle) and impulsive (salt and pepper) noise. To replicate Gaussian noise, we added a normally distributed random value to each pixel. For Salt and Pepper noise, we randomly altered pixels to extreme values (0 for dark, 1 for bright). Speckle noise, being multiplicative, was created using the formula out $=$ image$+n\times$image, with N as uniform noise defined by a specific mean and variance. In this study, we simulated these 3 noise types, as well as the isolated effects of salt (bright) and pepper (dark) values. Such a methodology enables a thorough assessment of the object detection algorithms' resilience against various noise types commonly found in satellite images.

Brightness & Saturation. Satellite images are influenced by various factors, including atmospheric conditions, optical properties, sensor characteristics and data processing techniques. For example, atmospheric scattering can reduce color intensity, resulting in lower saturation. Min addition, certain geographical areas, such as deserts and polar regions, can influence the image properties due to the high solar reflectance, which leads to increased

brightness [27](Figure 3.4.2 (d)). To replicate different brightness levels, we used a function from the OPENCV Python library parameterized with $\alpha \in [0, 1]$ and $\beta \in [-127, 127]$. We set α to 1 and varied β from 0 to 100 in steps of ten, with lower β values corresponding to lower brightness. To simulate different degrees of saturation, we used a function from the Pillow Python library with a single parameter. This saturation parameter ranged from 0 (colourless image) to 1.5 and was increased in steps of 0.15. These methodologies enabled us to evaluate the object detection algorithms' robustness against a spectrum of brightness and saturation levels, reflecting the diverse conditions encountered in real-world satellite imagery.

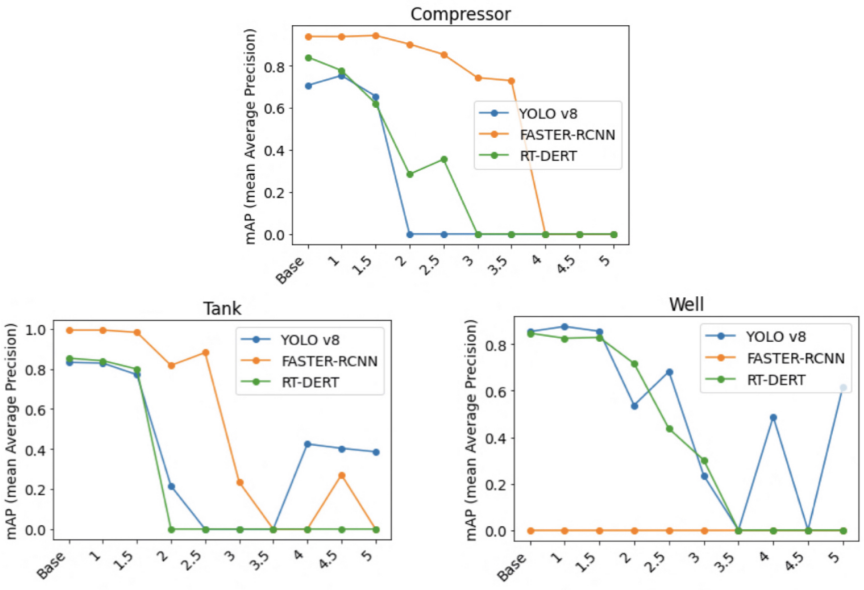


Figure 3.14: Impact of tangential resolution variations (from $0.5m$ to $5m$) on algorithms performances.

Results: Figure 3.14 illustrates the variations in mean Average Precision (mAP) for each object detection model as a function of spatial resolution changes, focusing on 3 types of infrastructures (compressors, tanks, and wells). We set the mAP to zero in instances where the models fail to detect and recognize the infrastructures. The effectiveness of the algorithms in recognising these objects varies with the resolution. YOLO v8 in particular can detect compressors at resolutions of up to $1.5m$, whereas FASTER-

RCNN extends this capability up to $4m$. For tanks and wells, YOLO v8 successfully detects and recognizes them at resolutions as high as $5m$, with mAPs fluctuating between 40% and 60%. It's important to mention that FASTER-RCNN consistently fails to detect wells, even in the base image. Furthermore, we observe that compressors are more challenging to detect at lower resolutions (detected up to $4m$), in contrast to tanks and wells, which are detectable up to $5m$.

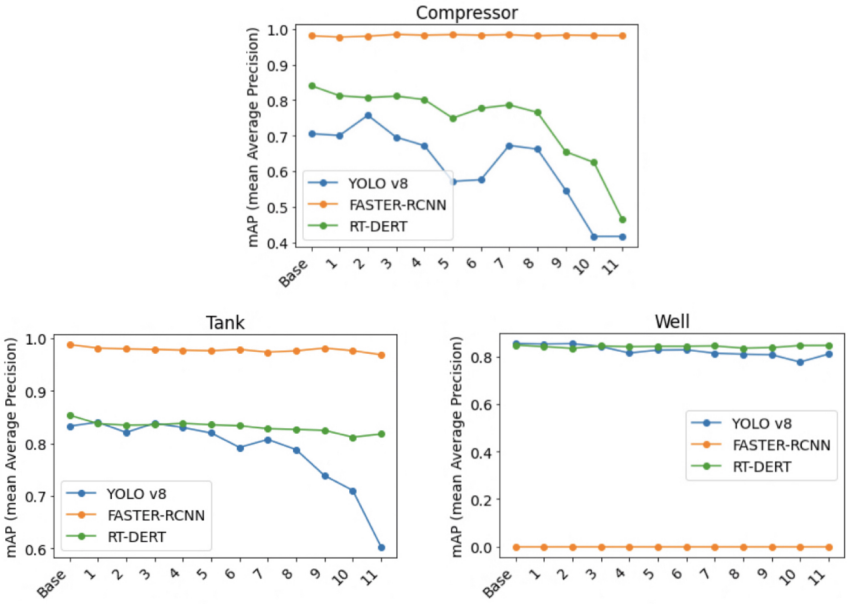


Figure 3.15: Impact of tangential distortion variations (left gradual stretching from level 1 to 11) on algorithms performances.

Figure 3.15 shows that FASTER-RCNN exhibits remarkable consistency in its performance, with minimal sensitivity to tangential distortions; the mean Average Precision (mAP) for compressors and tanks remains relatively stable across the different distortion levels. In contrast, YOLO v8 exhibits more significant fluctuations in mAP in response to these distortions. Both YOLO and RT-DETR show a notable decrease in performance at the 5th level of distortion when detecting and recognizing compressors. In the case of tank detection, YOLO is more adversely impacted than RT-DETR, particularly from the 7th level of distortion onward. Interestingly, both YOLO and RT-DETR maintain almost consistent performance in the tank case, unaffected by the levels of distortion presented.

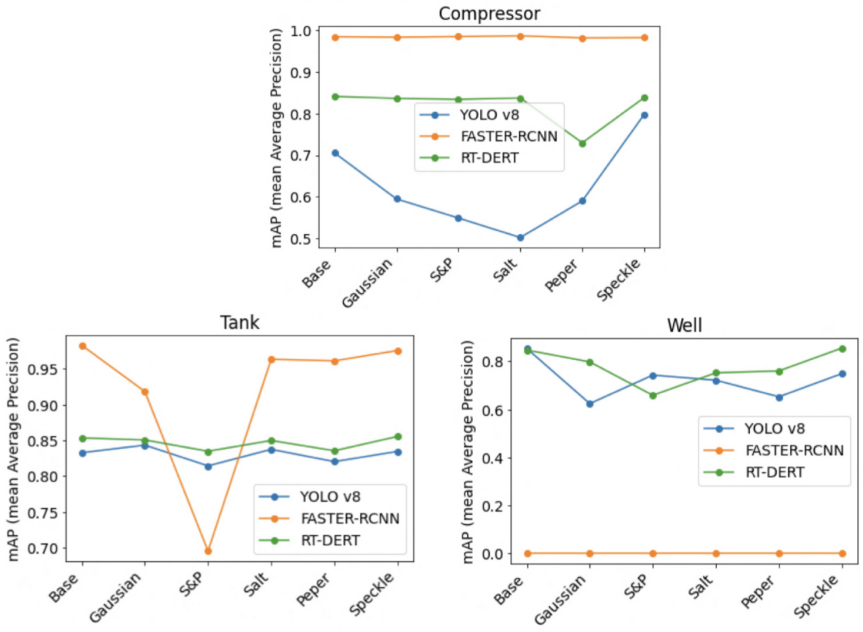


Figure 3.16: Impact of noises (gaussian, salt&pepper, salt, pepper and speckle) on algorithms performances.

Figure 3.16 shows the varying impact of different noise types on the detection of compressors. Notably, YOLO is significantly affected by salt noise, while RT-DETR is more susceptible to pepper noise. Intriguingly, YOLO demonstrates improved performance with speckle noise in detecting compressors, compared to its performance on the base image. For wells and tanks, RT-DETR exhibits greater resilience to the five types of noise than YOLO, with the notable exception of salt and pepper noise in well detection. On the other hand, FASTER-RCNN generally maintains stable performance, although it shows a heightened sensitivity to salt and pepper noise in the detection of tanks.

In Figure 3.17, we observe that increasing brightness levels (from 1 to 10) generally leads to a decline in the mean Average Precision (mAP) of the models. This effect is particularly evident in the case of compressors, where YOLO's performance begins to diminish from level 2. However, YOLO is less affected than RT-DETR when detecting wells. For compressors, both FASTER-RCNN and RT-DETR maintain a relatively steady mAP

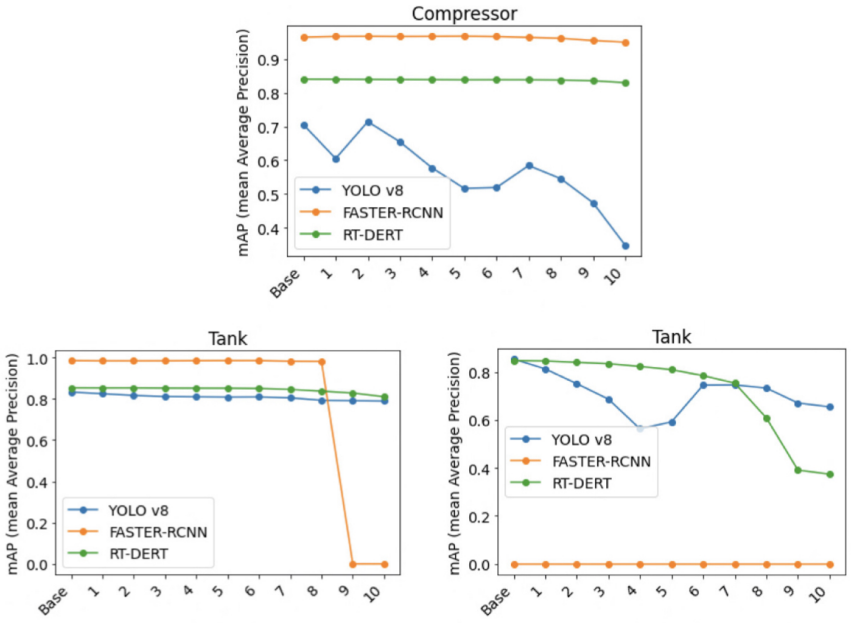


Figure 3.17: Impact of brightness on algorithms performances.

across the range of brightness levels. When detecting and recognising tanks, the performance of FASTER-RCNN drops significantly from level 9, which means that it can no longer perform detection.

Figure 3.18 shows, for the compressor case, that YOLO presents a higher mAP for images with low saturation, which decreases for higher levels of saturation. Conversely, FASTER-RCNN exhibits the opposite trend. Concerning RT-DETR, its mAP value remains almost constant over saturation variations for the compressor case. For the tank case, all 3 models seem to react similarly, with an increasing mAP over saturation augmentation. This observation holds true for the well case between YOLO and RT-DETR.

3.5 Discussion & Conclusion

To effectively mitigate methane emissions in the oil and gas (O&G) sector, a comprehensive emissions profile of each O&G infrastructure is essen-

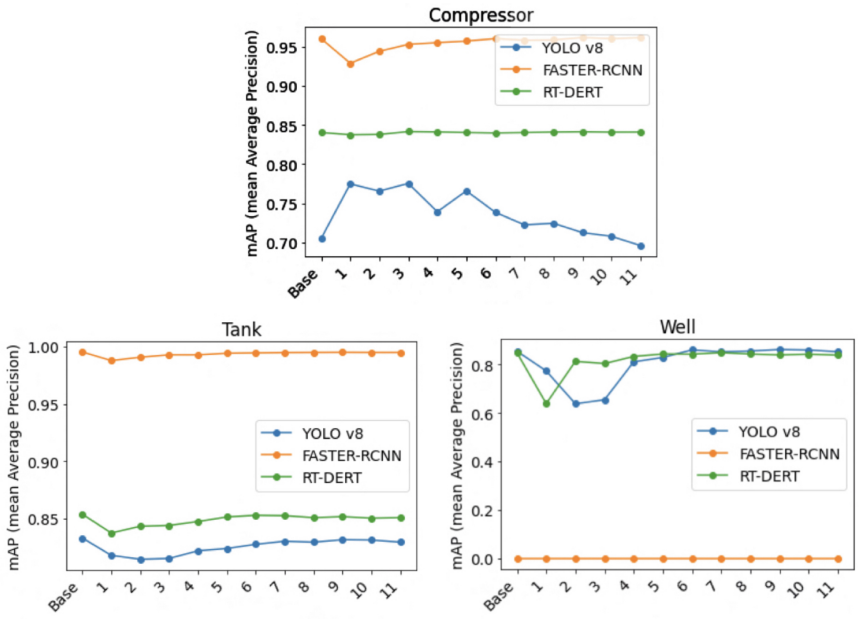
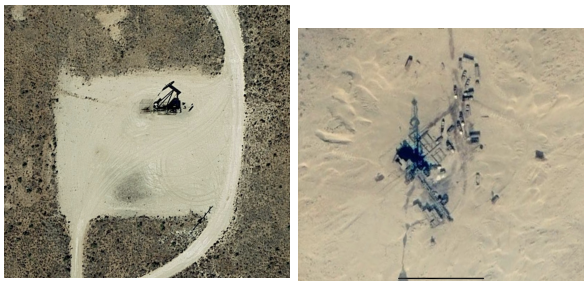


Figure 3.18: Impact of saturation on algorithms performances.

tial to enable for an in-depth understanding of individual emission trends. Therefore, satellite observed methane emissions from different oil and gas basins must be accurately associated with the particular infrastructures from which they emanate. This requires, as first step, an efficient solution to automatically detect and identify such infrastructures, a task that can be well addressed by detection algorithms. However, these algorithms come in myriad forms, each with a unique architecture and range of performance. While previous studies have compared specific algorithms for automatically detecting O&G infrastructures, these have typically focused on a single infrastructure type. In response to this limitation, our work case presents a comparative analysis of three main supervised algorithms-YOLO v8, FASTER-RCNN, and DETR-for the simultaneous detection of tanks, wells and compressors. A unique database of various aerial snapshots of O&G infrastructures in the Permian Basin, USA, was used for the study. The tests performed showed that YOLO v8 outperformed FASTER-RCNN and DETR in terms of accuracy, although it was not pre-trained. Interestingly, FASTER-RCNN showed a superior ability to detect compressor styles that are significantly underrepresented in our database (and in the Permian Basin) on certain occasions. We also conducted a systematic as-

sessment of the impact of several satellite image variations—including resolution changes, tangential distortion, noise, brightness, and saturation—on 3 algorithms. The experiment's results indicate that FASTER-RCNN was the least affected by these simulated variations, although it failed to detect any wells. Conversely, YOLO, despite achieving the highest mean Average Precision (mAP) post-training, exhibited the greatest sensitivity. The study underscores that the influence of satellite image variations on algorithm mAPs is highly dependent on the specific object being detected, highlighting the necessity of tailoring algorithm performance to the targeted objects. We identified a hierarchy in the impact of these variations on mAPs: resolution, noise, distortion, brightness, and saturation, in descending order of influence. This ranking offers valuable insights into their relative significance in affecting algorithm performance. By incorporating these impactful variations into the training dataset, our experimental approach seeks to bolster algorithm robustness in practical applications. The comparative analysis of the 3 algorithms sheds light on their individual strengths and weaknesses, providing crucial guidance for choosing the appropriate algorithm for specific tasks under varying satellite imaging conditions.



(a) Well Permian Basin, USA (b) Well Turkmenistan

Figure 3.19: Examples of different forms of well in USA and Turkmenistan O&G basins. *Source : @Google earth.*

Future works would benefit from investigating the reusability of this method in other O&G basins, and their adaptability to other learning paradigms. Indeed, from a country to another and even from a basin to another (Figure 3.5), O&G infrastructures could have a totally different design. To automatically detect and recognize new form of compressor, well and tank or a new infrastructure like flare it will be required to constitute other manual annotated database for the supervised approach. It would also be useful to extend this scope to "few-shot" and "self-supervised" algorithms

that require minimal or no training data in order to avoid time consuming manual annotation. Also, it would be interesting to investigate ensemble learning method in order to combines the individual advantages of our 3 tested models.

Table 3.1: State of the art papers and their used artificial intelligence method by categories.

	Method	Regression	Classification	Clustering	Dimension Reduction
Plume Detection					
[80]	CNN		x		
[135]	CNN		x		
[54]	GoogleLeNet,LSTM		x		
[116]	ANN		x		
[101]	CNN,SVM		x	x	
[129]	CNN		x	x	
Plume quantification					
[50]	CNN	x	x		
[133]	FC-LSTM,ConVLSTM		x		
[136]	CNN,ConVLSTM		x		
Plume Detection & Quantification					
[53]	ANN	x			
[109]	ANN,RSVM		x		
[8]	CNN		x		
O&G Infrastructure Recognition					
[105]	CNN		x		
[153]	CNN		x		
[34]	DBSCAN			x	
[37]	YOLO, ...		x		
O&G Equipment					
[59]	PCA,ANN	x			x
[138]	DBSCAN			x	
Emission Forecasting					
[134]	RF,..	x			
[52]	GBR,...	x	x		
[106]	CONVGRU	x			

Table 3.2: Pre-trained Algorithms Average Precision (AP) results in % on OG database. **The number of parameters is expressed in millions*

		Average Precision (AP)			
Model	**Parameters***	**Compressor**	**Tank**	**Well**	**Total**
		YOLO v8			
8m	25.9	**99.5**	**98.8**	79.4	**92.6**
8l	43.7	**99.5**	88.1	**80.3**	89.3
8x	68.2	98.8	90.9	73.6	87.8
		FASTER-RCNN			
R50-FPN	41.7	51.6	51.1	40.5	47.7
R101-FPN	60.6	**53.2**	**57.8**	35.4	**48.8**
R101-DC5	184.5	52.1	47.1	**42.9**	47.4
		DETR			
R50	41	94.9	75.1	72.7	80.9
R101	60	**100**	**80.4**	**77.1**	**85.8**
R101-DC5	60	91.9	69.9	67.9	76.3

Table 3.3: Non pre-trained algorithms Average Precision (AP) results in % on OG database. The empty spaces translate the non-convergence of the models and then the absence of results.

		Average Precision (AP)			
Model	**Parameters***	**Compressor**	**Tank**	**Well**	**Total**
		YOLO v8			
8m	25.9	97.2	**90.6**	**81.4**	**89.7**
8l	43.7	97.7	90.4	78.0	88.7
8x	68.2	**98.1**	87.2	77.2	87.5
		FASTER-RCNN			
R50-FPN	41.7	25.4	2.3	7.2	11.6
R101-FPN	60.6	—	—	—	—
R101-DC5	184.5	—	—	—	—
		DETR			
R50	41	—	—	—	—
R101	60	—	—	—	—
R101-DC5	60	—	—	—	—

Chapter 4

Automated Methane Plume attribution

This chapter delves into the sophisticated process of automatically associating satellite-detected point sources of methane emissions with specific sites, operators, and infrastructures previously cataloged. It showcases the integration of cutting-edge technology with existing data to improve the accuracy and efficiency of methane emission monitoring within the Oil & Gas (O&G) sector.

In Subsection 4.2, we introduce the O&G Profile framework, a novel system designed for the automatic association of satellite detections with operators and O&G sites identified during the PermianMAP ground survey. This innovative approach not only streamlines the process of linking satellite-detected methane emissions to specific operators and sites but also demonstrates the framework's capability to augment and extend the reach of ground surveys conducted at zero additional cost. Utilizing the PermianMAP as a case study, the O&G Profile framework exemplifies how integrating satellite data with ground-based observations can significantly enhance our understanding and monitoring of methane emissions over time.

Subsection 4.3 explores the application of K-dimensional trees algorithm for the automatic association of detected methane emissions with various O&G infrastructures. This method employs a spatial data structure to efficiently organize and query large datasets, enabling the precise match-

ing of methane plumes to the nearest infrastructures within the dataset. This approach significantly improves the speed and accuracy of attributing methane emissions to specific infrastructures, a critical step in addressing and mitigating the environmental impact of such emissions.

Finally, Section 4.4 addresses the inherent uncertainties associated with these multi-level associations between methane detections and their sources. It acknowledges the complexities involved in accurately identifying the origins of methane emissions and discusses the potential strategies for managing these uncertainties. This Section proposes a forward-looking perspective, suggesting the exploration of advanced statistical models, machine learning techniques, and continuous data refinement processes as potential avenues for enhancing the precision of methane emission attributions. Through these advancements, the chapter outlines a pathway towards more reliable and actionable insights into methane management and mitigation strategies in the O&G industry.

Contents

4.1 Introduction

The advancement of point source satellite sensor technology has significantly enhanced the precision of methane emission estimations, achieving levels of granularity that were previously unattainable. Furthermore, these innovations have expanded the capacity for quasi-continuous measurements, allowing for near-constant monitoring of methane emissions across vast geographic areas. These technological strides are set to revolutionize our approach to environmental monitoring by providing a wealth of detailed

data on methane emissions with unprecedented accuracy and temporal resolution.

As we move forward, the sheer volume of data generated by these sophisticated sensors presents both an opportunity and a challenge. The potential to transform this extensive dataset into accurate and actionable information is immense, yet it requires the development of equally advanced data processing and analysis methodologies. In this context, it is crucial to explore and implement cutting-edge data transformation techniques that can efficiently handle and interpret the massive amounts of data produced.

To address this challenge, future research and development efforts should focus on leveraging machine learning and artificial intelligence algorithms, which are adept at processing large datasets and extracting meaningful patterns. These technologies can be instrumental in automating the identification of methane emission hot-spots, tracking changes over time, and predicting future emission trends based on historical data. Furthermore, the integration of cloud computing and big data analytics platforms can provide the necessary infrastructure to process and store the vast quantities of data, making it accessible to scientists, policymakers, and industry stakeholders for informed decision-making.

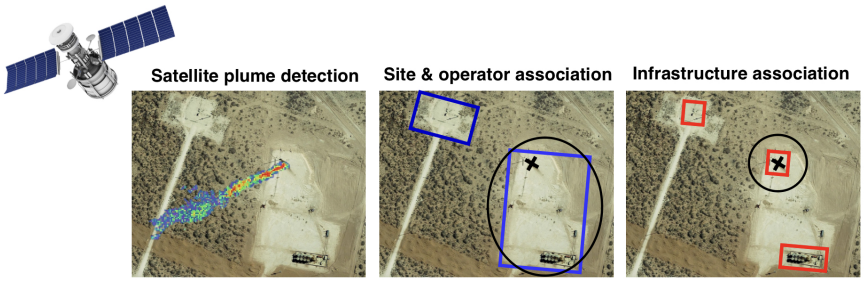


Figure 4.1: Detected Methane plume association to O&G site, operator and infrastructure by spatial location. *Source : @Google earth.*

The dataset derived from the fleet of ten satellites currently dedicated to methane emissions detection encompasses critical information about each detected methane plume. This includes the geographical coordinates (latitude and longitude) of the plume, the rate of methane emission (measured in kg/hr), and the associated uncertainty levels. Each methane leak detection, represented as a plume, is assigned a geographic location, which serves to pinpoint the potential source of the emission. However, the identification

of the geographic location of a detection within an Oil & Gas (O&G) basin is merely the initial step in a more complex analysis process.

Identifying the precise source of a methane leak requires more than just the geographical coordinates of the plume; it necessitates a comprehensive understanding of the types of sites, the specific infrastructures present, and the operators responsible for those assets. The mere geographical location of a detection does not inherently reveal the nature of the site (e.g., well, pipeline, storage facility) or the identity of the operating entity. This gap highlights the need for a sophisticated analytical framework that can integrate satellite data with detailed ground-level information about O&G operations.

To bridge this gap, integrating satellite detections with databases containing detailed records of O&G infrastructure types, their specific locations, and operator information becomes essential. Such integration enables a more nuanced analysis, facilitating the attribution of methane emissions to specific sources. Advanced algorithms and machine learning models can play a pivotal role in automating this process, analyzing spatial data to match methane plumes with potential sources based on proximity, operator activity, and infrastructure type.

To develop a comprehensive profile of methane emissions by site type, infrastructure, and operator, it is imperative to first establish the origins of each methane detection made by various satellite instruments. This involves determining the specific operator, site, and infrastructure associated with each detection, as illustrated in Figure 4.1. The task of associating methane emission detections from a wide array of instruments with a diverse range of operators, sites, and infrastructures across extensive areas such as entire Oil & Gas (O&G) basins can quickly become an overwhelming manual endeavor. Currently, there is a notable absence of automated methods for performing these associations, making the process a labor-intensive and time-consuming task that requires the expertise of specialists.

This manual approach to linking methane emission detections to their respective sources involves a meticulous examination of data and often relies on the expert knowledge of environmental scientists and industry professionals. They must sift through vast amounts of satellite imagery and emissions data, cross-referencing this information with detailed records of O&G infrastructure and operator activities. Such a process is not only slow but also prone to human error, limiting the efficiency and scalability of emissions monitoring efforts.

To unlock the full potential of satellite technologies for methane emissions monitoring in the O&G industry globally, there is a pressing need for the development and implementation of automated methods that can streamline the association of methane detections with their multilevel sources. Leveraging advancements in artificial intelligence (AI) and machine learning (ML), it is possible to create sophisticated algorithms capable of analyzing satellite data in conjunction with comprehensive databases of O&G infrastructure, site types, and operator information.

By automating the process, it becomes feasible to rapidly and accurately match methane detections with specific sources, thereby enhancing the efficiency and accuracy of emissions profiling. Such automation could significantly reduce the time and resources required for emissions monitoring, allowing for more frequent and comprehensive assessments. Moreover, automated systems can continuously learn and improve, adapting to new data and evolving industry landscapes to provide increasingly precise emissions insights.

In addition to improving the speed and accuracy of emissions monitoring, automated association techniques can facilitate a more dynamic response to methane leaks, enabling operators and regulatory bodies to take timely and informed actions to mitigate emissions. By integrating these automated methods into the existing frameworks for environmental monitoring and regulatory compliance, the O&G industry can make substantial strides toward reducing its environmental impact and advancing sustainability goals.

4.2 O&G sites and operators attribution

4.2.1 Context

The Permian Basin is the largest O&G basin in the United States, it covers 86,000 square miles of land across West Texas and Southeast New Mexico and is home to more to tens of thousands of O&G sites. Many works, as for example [152, 44, 14], have been devoted to the study of methane emissions over different angle of view, which make the Permian one of the most informed basins in terms of data. The Environmental Defense Fund (EDF), through the PermianMap project, has launched two GAO air campaigns and the collection of data from the CarbonMapper satellite to

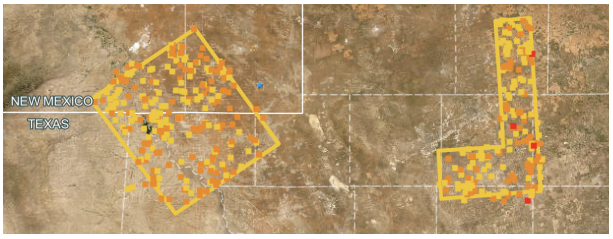


Figure 4.2: Perimeter of the study zone of PermianMAP survey (left: Delaware basin, right: Midland basin).

survey its emissions in the Permian basin. All these detections, according to their location, have been assigned to operators, type of site and emitting infrastructure. These open-source data allow indirectly to have access to the geographical coordinates of the sites and to their information concerning their type and the concerned operator. This information is particularly sought after and useful in order to be able to inform detections from any other satellite measurements. For this study case, we propose to use the data from CarbonMapper and GAO to label raw satellites point source detections in order to collect their emissions profile.

4.2.2 Datasets presentation

Point-source satellites characteristics. The point-source satellite current available data can be acquired via 8 satellites: Landsat-8, WorldView-3, Sentinel-2, GHGSat, PRISMA, EnMAP, EMIT and CarbonMapper. which are fine-pixel ($\leq$ 60m) instruments designed to quantify individual point sources by imaging of the plume. Their detection thresholds in the 100–10 000 kg/h range that enable monitoring of small to large point sources.[45] The point source satellite used in this study will be name Satellite X for confidentially reasons. Since the satellite X is a commercial program, its database are under non disclosure agreement, implying the impossibility for us to share information such as the location of the plumes detected and their methane rates. The results presented at the end of this chapter are a combination of data from the X satellite and the PermianMAP project.

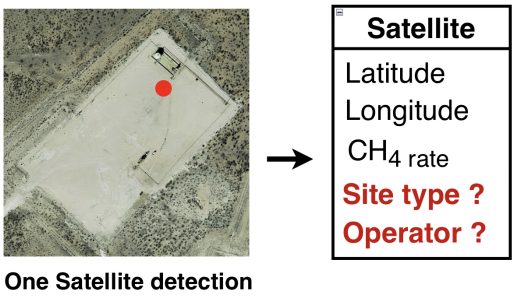


Figure 4.3: Illustration of satellite X informations for one detection. *Source : @Google earth.*

Satellite X dataset description Satellite X data consists of a set of detections of methane emissions from various sources around the world. To each detection is associated the estimated geographic coordinates of the detected plume (latitude and longitude)[1], the estimated amount of methane emitted and the uncertainty about this estimate (in %).This initial dataset covers the period 01/01/2021 to 30/09/2022.

PermianMAP project description. The Environmental Defense Fund launched the Permian Methane Analysis Project (PermianMAP), making near real-time methane monitoring data available to the public online, enabling oil and gas companies, regulators and other stakeholders to directly access this project measurements. The majority of these last ones are taken in a 10,000 square-kilometer grid that spans the Delaware Basin and a portion of Midland Basin, a subsection of the broader Permian Region. Methane plume detections have been observed with the AVIRIS-NG instrument installed on the Global Atmospheric Observatory (GAO) aircraft in coordination with the satellite CarbonMapper. The measurements were made during two campaigns in Summer and Fall 2021 on predetermined regions in both Delaware (5000 km2) and Midland (2500 km2) sub-basins. Flights occurred on all days with acceptable atmospheric conditions with the intent of mapping all predetermined regions at least three times during the flight window. All detected emitting sources were then manually inspected using a combination of high-resolution imagery collected by the aircraft and Google Earth satellite imagery to determine the segment of O&G production responsible: Production, Gathering & Boosting and Pro-

[1]It should be noted here that the uncertainty associated with the estimated location of each plume is not reported.

cessing. The observed plumes and emission sources were also attributed to the most likely responsible operators for each source by using a combination of open-source and commercial databases. Since the Permian area is very dense and the sites are not widely spaced, the attribution of an operator to a site contains uncertainty which is not quantified.

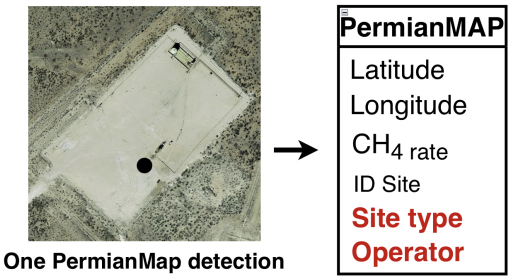


Figure 4.4: Illustration of PermianMap informations for one detection. *Source : @Google earth.*

PermianMAP dataset description. The selected data are distributed over two measurement periods, the first from 2021-07-26 to 2021-08-10 and the second from 2021-10-03 to 2021-10-17. These two campaigns have allowed the collection of 1696 detections in total. Each detection is associated with its geographical location (latitude and longitude), the estimated quantity of methane, the uncertainty of the estimation (in %), the site ID, the type of site, the infrastructure and the operator at the origin of this emission. The number of sites involved in PermianMap detections is known for each period of the campaign and can be obtained by counting the IDs of these sites. However, the detections obtained during the second period can come from already visited sites during the first campaign. So, the total number of visited sites from combined campaigns can not directly be obtained. Site ID of both the campaign are sequence of letter and number starting by "S2" for the Summer campaign and "F2" for the Fall campaign. Since we have no uncertainty in the determination of the location of methane plumes detected by satellite, we will propose an automatic association of plumes only at the site level and not at the infrastructure level which, due to their fine spatial granularity, requires more precision. It should also be noted that the absence of uncertainty measurements implies that a certain degree of unknown uncertainty is present in the association of our detections to given sites.

4.2.3 Study case objectives

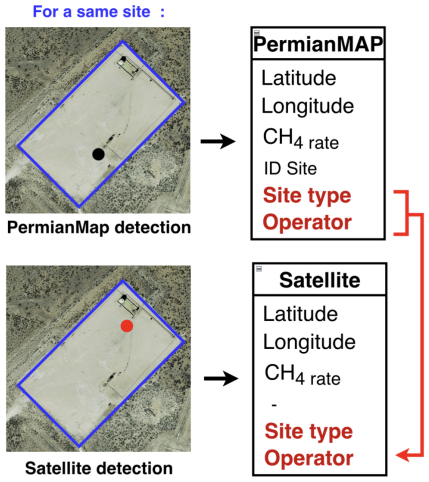


Figure 4.5: Representation of the transfer of information from PermianMAP to satellite X for a same site. *Source : @Google earth.*

The objective of our O&GProfile method is to provide an emission profile for each site but also to make it possible to associate satellite detections with the type of site and operator from which they come. Thus, all satellite detections can subsequently be studied independently of all other detections. In our case, the satellite X detections are initially not associated with the type of site or operator whence they come. PermianMAP detections are initially associated with the type of site and operators, so they will be used to label the satellite X data. However, in order for the PermianMAP information to be transferred to those of satellite X, it is necessary that they are on the same site in order to share the type of site and operator as illustrated in Figure 4.5. Once satellite X detections is labelled, it is possible to study the emission profile of a site from PermianMAP detections alone and/or from the combination of PermianMAP and satellite X detections. In the same way it is possible to draw up an emission profile by type of site and operator.

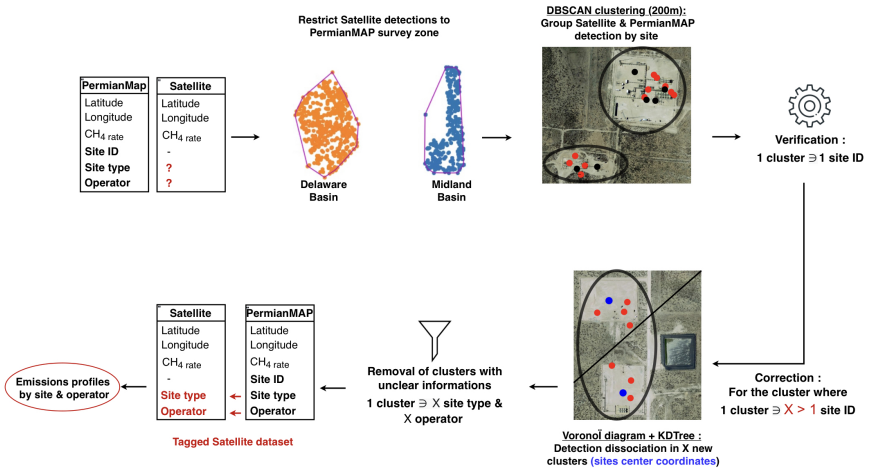


Figure 4.6: Presentation of O&GProfile method steps.

4.2.4 O&GProfile framework

In order to obtain emission profiles by site and operator composed of satellite X and PermianMap quantified detection, we propose the O&GProfile which is mainly based on the use of the DBSCAN clustering algorithm[25]. The O&GProfile consists of 4 main steps:

1. Selection of satellite X data included in the PermianMAP data study perimeter selection of satellite X data included in the PermianMAP data study perimeter;
2. clustering of satellite X and PermianMap detections (DBSCAN);
3. Verification and correction of the obtained clusters/sites using Voronoï diagram and Nearest Neighbor KDtree;
4. Transfer of information from PermianMap detections to satellite X detections when PermianMap and satellite X data are located on the same site.

In order to benefit from PermianMap information, satellite X detections must be included in the PermianMap study area. The PermianMap study area is divided into two sub-basins, the Delaware basin and the Midland

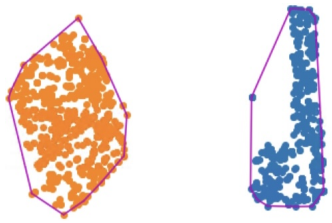


Figure 4.7: Representation of convex boundaries around 2 basins of PermianMAP study.

basin. In order to obtain the precise perimeter of each of them, the PermianMAP detections of the two basins were first separated via the use of the K-means[47] clustering algorithm.

The K-means clustering algorithm computes the centroïds and iterates until we it finds optimal centroïds. It is a parametric method which requires to know in advance the number of cluster k. In our case, the parameter k is fixed to k=2 in order to spatially dissociate 2 zones/basin (cluster 0 and 1). the detection points are then assigned to a cluster in such a manner that the sum of the squared distance between the data points and centroïd would be minimum. As output, each of the PermianMAP detection is now linked to one of the basin. For each of the two basins - which consist in two ensemble of detection - the ConvexHull function was applied to obtain convex polygon boundaries of each one. Finally, the geographical coordinates of the satellite X detections were filtered according to their belonging or not to each of the two basins boundaries.

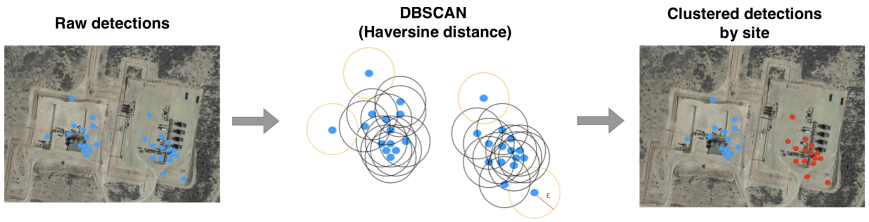


Figure 4.8: O&GProfile : DBSCAN clustering with harvesine distance. *Source : @Google earth.*

Concerning satellite X detections the number of sites involved is completely unknown (as no sites are referenced in these data), so an initial number of

sites from satellite X detection cannot be determined in advance. In order to obtain PermianMAP and satellite X detection grouped by site without a predetermined number of sites, spatial clustering method can be applied. The Density-based spatial clustering of applications with noise (DBSCAN) will be at the base of the O&GProfile method.

Contrary to other clustering methods (e.g. K-mean) DBSCAN is able to form clusters of arbitrary shape (not restricted to a spherical or convex shape), and can detect outliers, which makes it more suitable for our case where the data (detections) are spatially irregularly distributed over the sites. However, our study does not require the recognition and deletion of outliers, because each methane detection in our database represents important information about the quantity of methane emitted at a time t. DBSCAN works on the assumption that clusters are dense regions in space separated by regions of lower density. Indeed, by looking at the local density of the detections in large spatial dataset, DBSCAN groups densely grouped detections into a single cluster. The DBSCAN algorithm uses two main parameters:

- $minPts$: The minimal number of points to be present within the radius ε for the area to be considered "high density".
- Eps ε: A distance measure that will be used to locate the points in the neighborhood of any point.

The input data for the DBSCAN algorithm are the geographic coordinates of each detection (PermianMAP and satellite X): latitude and longitude. As used distance is geographic distance, the metric parameter used should be based on the haversine formula which determines the great-circle distance between two points on a sphere given their longitudes and latitudes:

$$haversin\left(\frac{d}{r}\right) = haversin(\phi 2 - \phi 1) + cos(\phi 1)cos(\phi 2)haversin(\lambda 2 - \lambda 1)) \tag{4.1}$$

where r as radius of earth, d as the distance between two points, $\phi 1$, $\phi 2$ is latitude of two points and $\lambda 1$, $\lambda 2$ is longitude of two points respectively

With the aim of grouping these detections by site, the Eps parameter will therefore be set at 200m which is the general length and width of

an O&G site. However, it will be necessary to fix the kilometers per radian fixed at $kms\_per\_rad = 6371.0088$ and to convert the 200m to radians: $0.2/kms\_per\_rad$.

The $MinPts$ is fixed at 1 a way to avoid that some detection can be considered by noise and then discard.

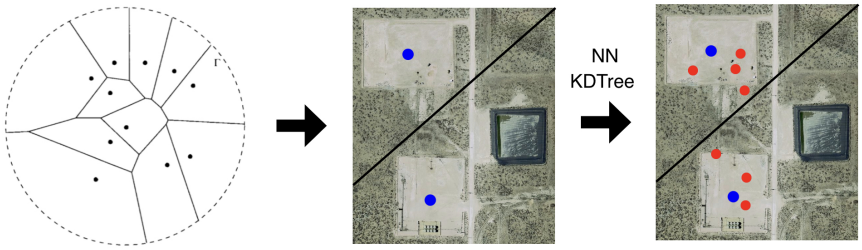


Figure 4.9: Two examples of detections (red dots) forming a single cluster spread over two sites centers (blue dots). *Source : @Google earth.*

4.2.5 Automated verification & corrections

Verification & correction : Voronoï diagram and KDtree algorithm Recall here that to transfer PermianMAP site type and operator information to satellite X detections it is required that they are positioned on the same site (cf. Figure 4.10). Therefore, clusters comprising PermianMap and satellite X measurements will be targeted in this step, and it will be verified if each of these clusters corresponds to an O&G site.

The verification process consists in counting site IDs (from PermianMap detections) by cluster, 3 scenarios are possible here:

- If an ID is found in several clusters, it means that the *Eps* is too small and that a site is divided between several clusters as represented on the left image of the Figure4.10;
- If a cluster contains several ID sites "S2" and/or "F2" it means that

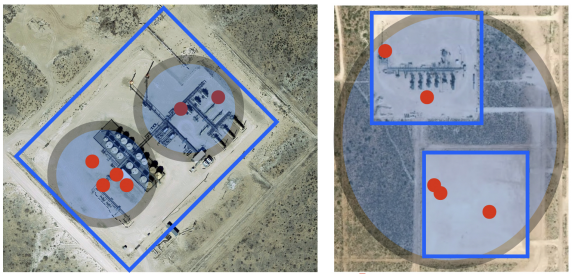


Figure 4.10: Example of clustering errors: (left) *Eps* too small, site break in two clusters (right) *Eps* too big, 2 sites grouped in one cluster. *Source : @Google earth.*

the *Eps* fixed at 200m is too large and that the cluster gathers several sites as illustrated on the right image of the Figure4.10;

- In the case where a cluster contains 1 unique ID then the cluster represents a site.

The last point represent the desired output, but in the case where the two others situations describe on the Figure 4.10 are identified then a correction process is require in order to correctly attribute detections to one site by cluster.

The correction procedure is based on the use of a Voronoï[2] diagram and the Nearest Neighbor (NN)[21] KDtree algorithm. The Voronoï diagram is a system created by scattering points at random on a Euclidean plane. The plane is then divided into tessellating polygons, known as cells (one around each point) consisting of the region of the plane closest to that point. The segments in a Voronoï Tessellation correspond to all points in the plane equidistant to the two nearest sites. The Voronoï diagram, taking as input the latitude and the longitude of the centers of sites established previously, permits to represent the possible division of the space according to the centers of site as represented on the Figure 4.9. The second step of the correction consists in linking the detections of each cluster to the centers of the two closest sites in order to constitute two new clusters for each site as illustrated on the last image of Figure 4.9. For this purpose, the nearest neighbor (NN) algorithm with the multi-dimensional search algorithm K-dimensional tree (KDTree) is used. The NN problem formulated here is to find the set of detections P_c in a data set P that are "closest" to a

query point q (site center), as measured by a distance function $d(p, q)$. The KDTree is used partitioning spatial points using a tree-like data structure that is useful for searching for the nearest point or nearest neighbors to a given point.

4.2.6 Results

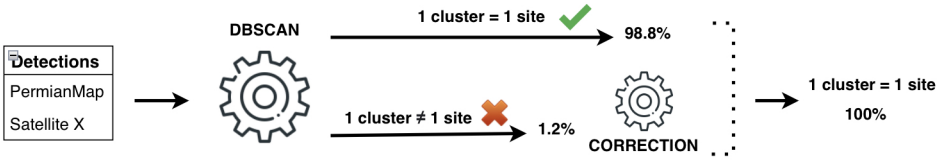


Figure 4.11: O&GProfile results in % of correct association for each step.

After the matching process, it was found that 103 satellite X detections belonging to the PermianMap study area spread over the two basins. These satellite X detections are added to the PermianMap detections and constitute a total dataset of 1799 detections. The location (latitude and longitude) of each this detection database is then insert as input to the DBSCAN algorithm which permits to obtain 879 clusters. Each of these cluster is supposed to represent a single O&G site. Among these 879 clusters, there are :

- 793 cluster only containing PermianMAP detections;
- 32 cluster containing PermianMAP and satellite X detections;
- 54 clusters only containing satellite X detections.

Since satellite X detections must be co-located with PermianMAP data in order to be labeled and included in the study, the 54 clusters with only satellite X detections will be removed from the study. It will be considered therefore that 825 clusters (783 + 32).

After the verification procedure, the case where the fixed *Eps* does not seem to be too small (left picture of the Figure 4.10) was not found On the other hand, the second case where the *Eps* would be too large was detected on 21 clusters so 2,5% (21/825) of the initial clusters (right picture of Figure

4.10). In order to understand the origin of these errors, the 21 clusters were visually and manually inspected revealing that 11 of them (1,3% (11/825) of the initial cluster number) have anomalies related to the site IDs and 10 others (1,2%) directly related to the established *Eps*.

These 11 clusters with detected anomalies contains each several site IDs for a same and unique site. This error of annotation from the PermianMAP database implied that several IDs are attributed to each of these 11 sites. These 11 clusters will not be submitted to the correction process has the error not comes from the DBSCAN clustering. Concerning the 10 other clusters containing 2 sites each, a correction must be applied in order to dissociate the intra-cluster sites. The itinerary of this procedure starts with the extraction (manually) of the coordinates (latitude and longitude) of the centers of each site that the clusters contain. At the end of the correction process 10 anomalous clusters were replaced by 20 new clusters where each one represents a single site. The number of clusters(=sites) after this process is 835 (825-10+20).

Once the detections are correctly distributed by site, the membership of a single operator per site must also be checked. Note here that in the PermianMap study, the manual association of sites to operators contains a certain amount of uncertainty that can lead to association errors. After verification, 12 clusters (1,4% (12/835) of the initial clusters) contain multiple operators, only populated from the PermianMAP measure. These last one have been deleted not to introduce confusion by the presence of several operator in only one site which is not possible. The final number of obtained cluster is of 823 clusters composed of 32 sites containing PermianMap and satellite X detections and 791 sites only containing PermianMap detections.

As summarized in the Figure 4.11, the DBSCAN algorithm allowed the grouping of 835 clusters where 1.2% (10/835) of these clusters presented an error and the other 98.8% were correctly grouped. The correction process has enabled to rectify the 1,2% of errors and thus obtain in 100% of the cases a cluster corresponding to a site.

Thanks to the O&GProfile method, the raw satellite X detections can now be interpreted independently (Table 4.2 of the PermianMap data (Table 4.1 according to the type of site and the operator at their origin. For sites/operators where both satellite and PermianMap measurements are available, the detections data can be gathered to form a time series concerning the quantity of methane emissions in time for a precise site/operator

Table 4.1: PermianMap data : number of plume and average CH4 k/hr by site type.

PermianMap

Site type	Number of plume	Avg CH4 (kg/h)
Production	446	372.04
Gathering & Boosting	494	423.06
Processing	74	514.09

Table 4.2: Satellite X data : number of plume and average CH4 k/hr by site type.

Satellite X

Site type	Number of plume	Avg CH4 (kg/h)
Production	8	712.45
Gathering & Boosting	13	2061.98
Processing	12	1285.84

(=site emission profile). It is also possible to directly compare the emissions quantities levels of PermianMap and satellite X instruments.

4.3 O&G infrastructure attribution

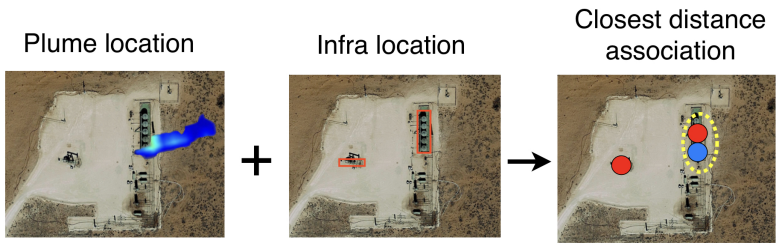


Figure 4.12: Methane Plume and Infrastructure association. *Source : @Google earth.*

The object detection algorithms presented in Section 3.2 are used to obtain, for each infrastructure detected and recognized, the coordinates of the bounding box delimiting the infrastructure in question and its type (tank, well, compressor). In order to associate a detected methane plume with the

infrastructure at its origin, we assume that the infrastructure at the origin of the methane emission is the (spatially) closest one to the estimated location of the methane plume as illustrated on the Figure 4.12.

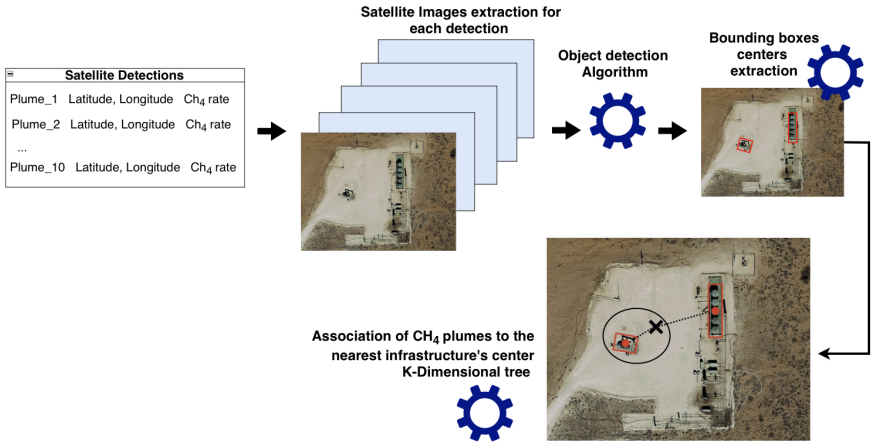


Figure 4.13: Methane plume attribution to infrastructure at its origin by the shortest haversine distance using K-Dimensional Tree. *Source : @Google earth.*

The framework shown in Figure 4.13 consists of several sub-steps for associating methane detections with the infrastructures that produced them. A dozen methane detections were first selected from the Permian basin. At the location (latitude, longitude) of each of these detections, a satellite image was extracted using Google Earth to visualise the surrounding infrastructures. The YOLO v8 model was run on these images in order to detect and recognise the infrastructures present on each image. Once the bounding boxes had been obtained for each infrastructure detected, the coordinates of the centre of these bounding boxes were extracted. Similarly to the O&GProfile semi-automated correction method presented in the previous Subsection, the Nearest Neighbor (NN) KDtree algorithm based on haversine distance was used. But instead of looking for k nearest neighbors the search only focuses on the first nearest neighbor. Grouping each methane detection to an infrastructure makes it possible to link to each detection the infrastructure at its origin

4.4 Discussion & conclusion

O&G Site and Operator attribution. The O&GProfile, an automatic method based on clustering methods, is the first method enabling the grouping of detections by site for the study of emission profiles by type of site and operators. It makes it possible to automatically label raw satellite detections in order to independently study and compare detections from different instruments. The O&GProfile has been tested in the Permian Basin (specifically the Midland and Delaware Basins) with labeled detections from the EDF PermianMAP project (CarbonMapper and GAO) and unlabeled satellite X detections. The association of these detections to sites via the use of the DBSCAN algorithm achieved a success rate of 98.8%. The 1.2% of errors was corrected with the Vornoï-NN-KDTree algorithm, resulting in a success rate of 100%.

Thanks to our O&GProfile, emission profiles for the period 2021-2022 have been obtained from 823 different O&G sites. Each site corresponding to a type of site (Production, Gathering & Boosting, Processing), it was also possible to establish emission profiles by site. This was also applicable for the operators of each site. The O&GProfile has also allowed the labeling of 103 satellite X detections divided into 33 clusters, which allows for independent analysis of satellite X data and/or comparison with PermianMAP data. Indirectly, our O&GProfile method has permitted to extend in time the PermianMAP study conducted from 2019 to 2021 by attaching to each surveyed sites new measurements from satellite X from 2021 to 2022 at zero cost. The O&GProfile method could easily be applied to different unlabeled satellite , while the ground infrastructure data (type of site and operators) can also be easily replaced by others data (most complete and/or in other regions). The process of correction still contains a part of visual and manual inspection concerning 1,2% of the cases, where site have to be distinct by extracting their middle geographic position. This step could be replaced in future studies by automated methods able to dissociate sites by images analysis and then automatically extract for each of them their middle geographic position.

However, our O&GProfile method relies on up-to-date O&G surveys/inventory, incorporating the geographic coordinates of all existing O&G infrastructures. Currently, several datasets in free access are available:

- Globally, datasets such as GOGi[98][2], HIFLD[3], GHGRP[4], and EIA cover the entire U.S.
- The Oil and Gas Infrastructure Mapping database (OGIM, [95]) provides detailed information, particularly for the conterminous US.
- At the state level, resources include the Texas Commission on Environmental Quality (TCEQ)[5] and the New Mexico Environment Department (NMED[6]).
- At the basin level, the Permian Methane Analysis Project (PermianMAP)[7] from the Environmental Defense Fund offers valuable insights.

Despite the detail provided by these datasets, they lack systematic updates and often represent point-in-time surveys (e.g., PermianMAP includes airborne data for 2019 and 2021). Given the rapid evolution of the oil and gas landscape, these "frozen" surveys become quickly obsolete. Object detection algorithms presented in the Section 3 could respond to this limitation by automatically detect O&G site type, but determining the sites' operators remains dependant on available informations.

O&G Infrastructures attribution. In order to attribute an infrastructure to an emission, we assumed that the infrastructure at the source of the emission was the one closest to the location of the emission. However, the location of methane emissions remains an estimate with a degree of uncertainty (cf. Subsection 2.2.2). This uncertainty can make the allocation process complex.

Indeed, as illustrated on the Figure 4.14, there is an obvious attribution case and a complex one. Evident cases make reference to site representation where infrastructures are alone or separated with large distance. The complex cases make reference to site where the space between infrastructures is very thin (e.g. less than 10m). In complex case, uncertainty concerning plume location could, even in small percentage, could lead to wrong associations. Point source satellite like satellite X enables to obtain a good precision concerning plume location (sufficient for site association), but could potentially lead to wrong infrastructures association.

[2]https://edx.netl.doe.gov/group/global-oil-gas-infrastructure-gogi-group
[3]https://hifld-geoplatform.opendata.arcgis.com
[4]https://www.epa.gov/ghgreporting
[5]https://www.tceq.texas.gov/toxicology/q-a/natural-gas
[6]https://gis.web.env.nm.gov/oem/?map=methane
[7]https://www.permianmap.org

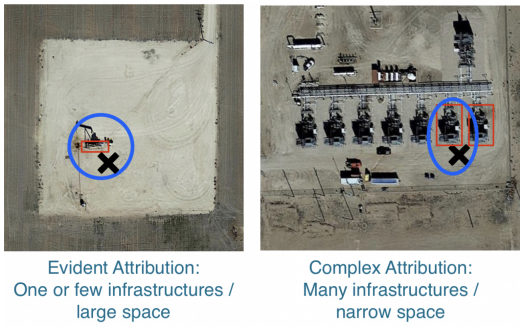


Figure 4.14: Representation of evident (left) and complex (right) infrastructures attribution depending on O&G sites configuration. *Source: @Google earth.*

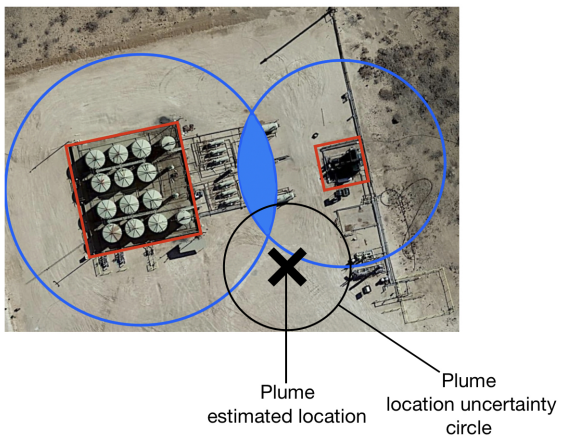


Figure 4.15: Plume attribution to infrastructures in complex case according to plume location uncertainty. *Source : @Google earth.*

A possible solution could be to use methane plume estimated locations uncertainty to determine a perimeter of uncertainty around each detection. This perimeter could be used to determine which infrastructure has he highest probability to be at the origin of the detection (Figure 4.15). Indeed, as plume location contains uncertainty, the output of the association process should be express in probability of belonging to infrastructures around.

Chapter 5

Dynamic and Intelligent Methane Emissions Inventory

This chapter introduces our current works and perspectives for the Dynamic and Intelligent Methane Emissions Inventory (DIMEI) framework, a comprehensive system that combines object detection (discussed in Chapter 3) and automatic association methodologies (explored in Chapter 4) for enhanced monitoring and inventory of methane emissions.

Section 5.1 details the structure and operational mechanisms of the DIMEI framework, emphasizing its dynamic aspect for near real-time monitoring (work in progress) and its intelligent component for emissions profile determination and forecasting (perspectives). This Section, based on our perspective publication of DIMEI [35], details how DIMEI could leverages cutting-edge technology to identify, classify, and monitor methane emissions from Oil & Gas (O&G) infrastructures.

In Section 5.2, we explore the potential integration of the DIMEI inventory with existing and future methane mitigation regulations. This part examines how the futures data and insights generated by DIMEI could inform policy development, facilitate compliance monitoring, and enhance the effectiveness of regulatory frameworks aimed at reducing methane emissions.

It discusses the framework's capacity to provide regulators and stakeholders with actionable intelligence, thereby enabling more targeted and efficient mitigation strategies.

Finally, Section 5.3 presents the prospective avenues for the implementation and expansion of the DIMEI framework. It outlines potential enhancements to the system's algorithms and data acquisition methods, aiming to broaden its application beyond the initial focus areas. The Section also considers the implications of integrating new technologies and methodologies, such as advanced deep learning models and satellite data with higher resolution and frequency, to improve the framework's accuracy and applicability. Furthermore, it discusses strategies for extending DIMEI's reach to cover more geographic areas and additional types of O&G infrastructure, thereby increasing its utility and impact in the global effort to mitigate methane emissions.

Contents

5.1 DIMEI Framework Presentation

5.1.1 General

The Dynamic and Intelligent Methane Emissions Inventory (DIMEI), illustrated in Figure 5.1, is an advanced framework that synergizes object detection capabilities with automatic association techniques experimented in previous Chapters 3 & 4. This innovative approach enables the detection and recognition of Oil & Gas (O&G) infrastructures and the subsequent association of methane plumes to their originating sources. Specifically, DIMEI focuses on profiling methane emissions from three primary types of

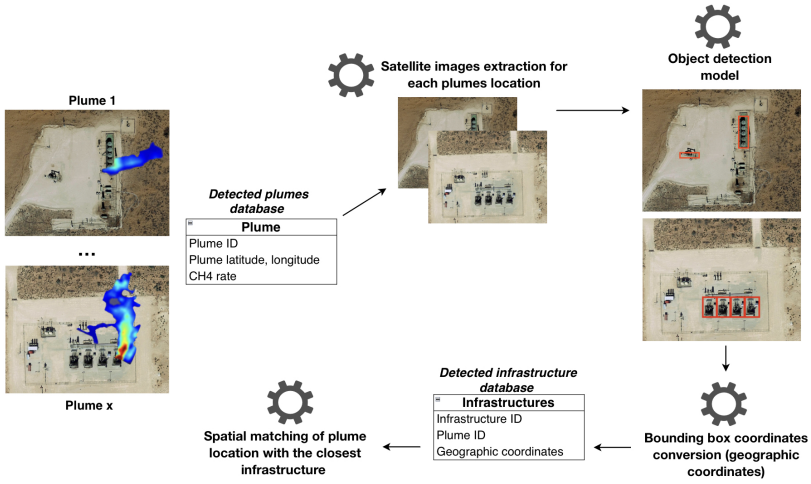


Figure 5.1: Presentation of the dynamic methane inventory framework. *Source : @Google earth.*

O&G infrastructures —wells, tanks, and compressors— within the Permian Basin.

At the core of DIMEI is a comprehensive database that is based on point source satellite methane emissions detections. This database encompasses critical information, including the estimated geographical coordinates (latitude and longitude) of each detected methane plume and its emission rate (measured in kg/hr). This repository serves as the operational foundation from which the framework initiates its analytical processes.

Indeed, upon the detection of a methane plume at a specific location, DIMEI triggers the extraction of a high-resolution satellite image snapshot corresponding to the identified geographical position. We precise here that the current version of our DIMEI do not integrates an API for automated snapshot extraction, extractions are therefore carried out manually using Google Earth. Then, each snapshot is subsequently processed by an object detection algorithm (DIMEI first component), which is meticulously designed to automatically identify and classify the type of O&G infrastructure depicted in the image. As a result, each processed snapshot yields a detailed list of detected infrastructures, each tagged with a label and precise geographical coordinates, thereby facilitating an accurate and

comprehensive profiling of methane emissions.

The second component consists in matching the infrastructure closest to that of the detected plume, using K-nearest-neighbor algorithm. Once the distances between a methane plume and surrounding infrastructures have been determined, the infrastructure with the shortest distance is considered as the infrastructure at the plume origin. Thus, for each plume the type of the closest infrastructure is attributed as its source. The final obtain database, is composed of x rows corresponding to the x detected plume from a high-resolution point source satellite, and each row contains the following variables :

- Plume ID
- Plume Latitude, Longitude
- CH4 rate
- Infrastructure ID
- Infrastructure type

Carried out successively, this method allows to build up a time series of the level and frequency of methane emissions by O&G infrastructures. This framework, thanks to the object detection component, has the advantage to offer instantiated detection of potential methane-emitting infrastructures, without recourse to fixed databases/inventories of oil and gas infrastructures. Indeed, the landscape of O&G infrastructures is constantly evolving, and the use of an inventory database produced at time t can quickly become inaccurate. In the context of quasi-continuous monitoring, the principle of snapshot instantiation is essential for building up an up-to-date inventory of infrastructures (provided to have access to regularly updated satellite images).

Expanding upon this foundation, DIMEI's methodology could not only enhances the accuracy of emissions tracking but also significantly contributes to our understanding of methane emission patterns across different types of infrastructures. By leveraging advanced machine learning techniques and high-quality satellite imagery, DIMEI stands at the forefront of environmental monitoring, offering a powerful tool for researchers, policymakers, and industry stakeholders alike. Furthermore, this framework's modular

design allows for future enhancements, including the integration of additional types of infrastructures and the expansion into other regions beyond the Permian Basin, thereby broadening its applicability and impact in the global effort to mitigate methane emissions.

5.1.2 Dynamic aspect

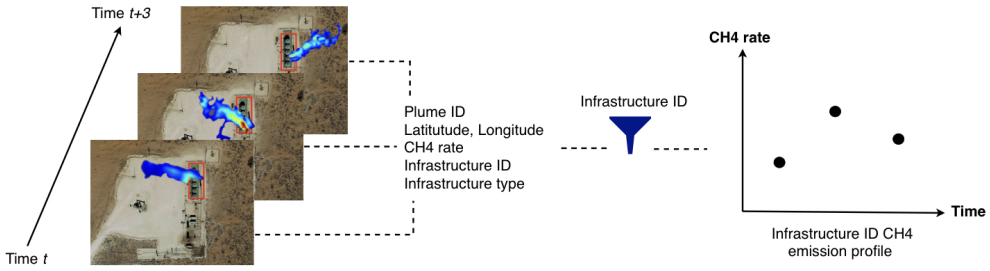


Figure 5.2: Inventory dynamic aspect : emissions profile determination. *Source : @Google earth.*

As illustrated on the Figure 5.2, the iteration of DIMEI over time allows to acquire temporal and spatial informations about methane emissions for each of 3 oil and gas infrastructures. Gathering this information over time enables to collect for each single infrastructure its emission frequencies and levels over time, thus constituting emission profiles (time series).

Using satellite detections of the X satellite source point in the Permian Basin, the DIMEI framework presented in Figure 5.1 enabled us to understand the dynamic aspect of the latter. Over the period 01/01/2021 - 30/09/2022, DIMEI enabled us to obtain multiple emissions for the same infrastructure.

As the current version of the DIMEI framework does not include an API for automatic snapshot extraction, it would have been time-consuming to extract manually a snapshot for more than 1000 methane plume detections. For this reason, we first selected detections for two O&G sites, where for each site several detections are available. After running the DIMEI framework, we obtained a first emission profile comprising 3 methane detections from a well (site 1) and a second emission profile comprising 5 compressor detections (site 2). As illustrated on the Figure 5.2 these emission profiles

give access to the temporal frequency of methane emissions and their rate for each of the two infrastructures. We are currently continuing this work in order to extend the study to various other sites. We also plan to integrate detections data from satellite X and others point source satellites in order to obtain longer time series in order to obtain a better characterization of infrastructures emission behaviors.

Collected over a sufficiently long period, these data will represent the base used for the intelligence part of the inventory, enabling to make spatio-temporal forecasts of methane emission levels for each infrastructure and group of infrastructures. The presented framework enabled to identify the recurrence of methane emissions over time for a number of specific infrastructures, thus characterizing the dynamic aspect of our framework.

5.1.3 Intelligent aspect

By being in possession of methane emissions profiles (spatio-temporal time series) of each for each O&G infrastructure updated in near real time, these profile can be aggregated by levels site/basin/region/country and even by company (at the condition to have access to company's site location).

The knowledge of the behavior of methane emissions from a specific infrastructures or level is mainly based on the characterization of the fluctuations. Time series at all levels can be analyzed in order to reveal three major diagnostic variables:

- Trends : Methane emissions long-term increase or decrease
- Seasonality : variations due to market demands and production rates
- Irregular variations in methane emissions quantity :
  - Episodic - unpredictable but identifiable
  - Random / Residual variations : unpredictable and uncontrollable in nature.

In order to precisely characterize the oscillations of the methane emissions, a regression can be established in order to determine the factors behind these oscillations as presented on the Figure 5.3. The methane emission

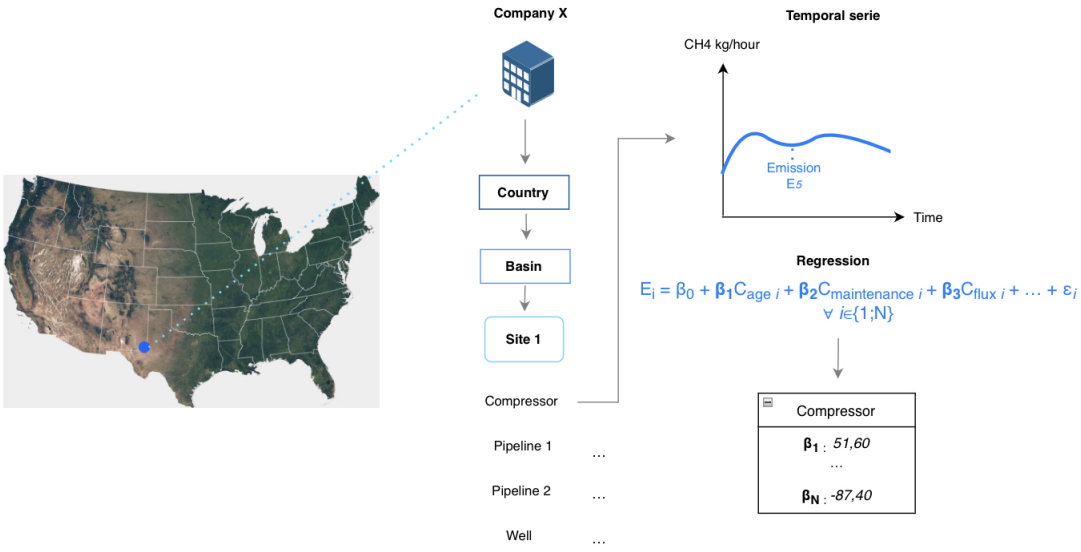


Figure 5.3: DIMEI general representation.

quantities $E_i \forall i \epsilon \{1 : N\}$ represent the dependant variable in the equation, as for the the explanatory variables they represent all compressor components determined to have an influence on the amount of compressor emissions. The regression coefficients $\beta_{\{1:N\}}$ associated with the explanatory variables will be obtained and will allow to interpret in a quantitative way the effect of each explanatory variable on the quantity of emissions. Note that the values of $\beta_{\{1:N\}}$ can be negative or positive and analyzed statistically (e.g., elasticity coefficient, marginal effect)

The values of the $\beta_{\{1:N\}}$ coefficients can also be collected as a time series to establish trends in the impacts of the explanatory variables, exploring the causes of the variations in methane emissions by company, by country, and by infrastructure type. Combining these data with the infrastructure data in the comparison framework will allow for example to identify the differences in emissions between compressors of two different companies, or to evaluate the quality of the equipment and maintenance between different O&G groups.

These analyses will benefit from integrating additional machine learning methods. The Figure 5.4 shows the phenomenon of irregular variations in

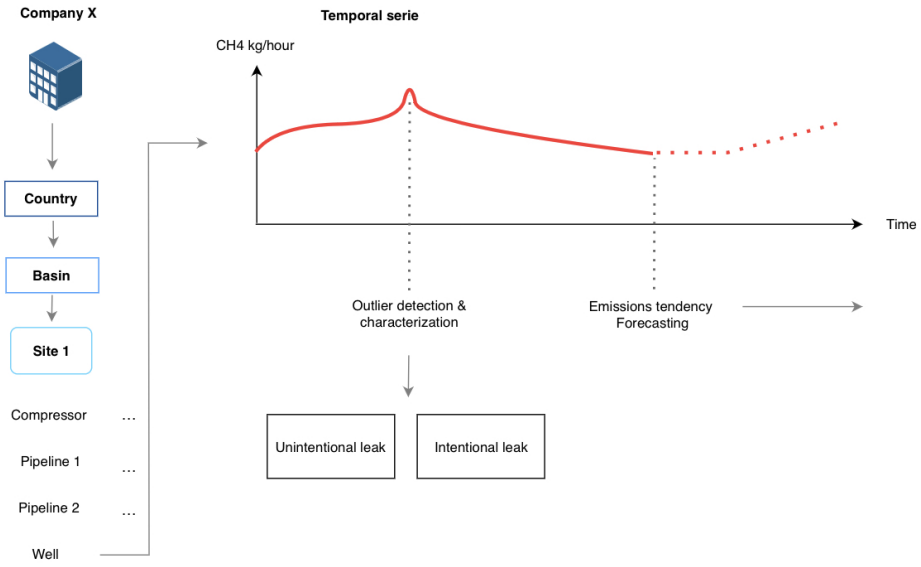


Figure 5.4: Presentation of the DIMEI intelligent aspect.

emissions at a given well. The above analyses allow us to determine the coefficients of the explanatory variables for an increase or decrease in methane emissions. However, if a leak is not related to the operating process of a facility, the explanatory variables are not representative of an increase in the amount of emissions; it then can be assumed that the nature of the leak is voluntary because it is different from the characteristics of involuntary leaks. This type of unusual behavior can be detected automatically through the use of anomaly detection algorithms that will allow automatically dissociate the cases of voluntary and involuntary leaks. Moreover, the accumulation of data related to voluntary leaks will allow to draw up the statistical profile and to characterize them better.

It is also possible to incorporate a predictive capability such as predicting methane emissions in the coming months to years if the historical time series is sufficiently constrained by data. The predictive aspect allows to create an essential margin of action for the mitigation of methane emissions. For example, it is possible to estimate the monthly emission quantities of a proposed new O&G site project based on the knowledge of the company's general emission profile, sites and facilities. Prediction of potential

breakdowns (predictive maintenance), simulations for optimal site/facility working process which reduce emissions, are also potential outcomes of such a system. The occurrence of voluntary and involuntary leakage on the basis of their characterization is also within reach. These predictions can be based on the use of statistical tools such as inferences or machine learning approaches (classifications, regression and generative models). The presented learning, analysis and forecasting methods are mainly based on the operation of O&G sites, hence linked to regulation factors, such as the evaluation of the influence of a regulation on the amount of methane emitted, or even to establish the automatic determination of the optimal regulation in relation to the amount of methane emitted and the behavior/practices of the oil groups.

5.2 Mitigation Regulations and DIMEI symbiosis

5.2.1 Regulation role for DIMEI

One of the main requirements for the implementation of the DIMEI inventory is the acquisition of activity and emission data from companies. The regulations of each country or state play an important role in the acquisition of this data by encouraging or requiring companies to maintain records and reports on their measurement campaigns and estimates. These requirements already exist within the government as part of the jurisdiction of the UNFCCC national inventory program. However, in some cases, these requirements must be revisited to ensure they adequately support national methane regulations and goals.

Based on this, the OGMP - first multi-stakeholder partnership - works on methane emissions reporting by providing a protocol to help companies systematically mitigate their methane emissions from O&G operations. The OGMP Reporting Framework is the highest standard of methane reporting, requiring companies to report methane emissions from all sources at both operated and non-operated ventures across the O&G value chain at an unprecedented level of accuracy and granularity.

The OGMP framework is divided into several reporting levels, including the "Gold Standard Reporting": a set of procedures for an empirical rec-

onciliation of measurements at source (Level 4) and site (Level 5) level. This "Gold Standard Reporting" is the central element of the DIMEI inventory allowing to acquire the most accurate emission estimate possible by taking advantage of complementary ground and air measurements. This OGMP initiative is based on the voluntary action of companies to join this movement, but it can nevertheless inspire a model for the imposition of the same procedure at the state, national, or regional level. In fact, the latest version from the European Council for a new methane regulation (published on December 2022) uses the OGMP 2.0 Framework as a benchmark for methane emissions reporting and requires the quantification of site-level methane emissions for all operated assets. According to IEA, once these informations are locally or nationally collected, the local or national jurisdiction will need a mechanism to verify its accuracy. This may include direct verification through inspections or third-party measurement. Or companies can be based to certify their compliance with regulations and submit independent audits in their submission.

More generally, after acquiring company reports and all other emission data from ground or air measurements, a common database must be developed to gather all these measurements and make them publicly accessible. For this purpose, The IMEO initiative, launched at the G20 Summit by UNEP, is in charge of collecting and integrating diverse methane emissions data to establish a global public record of empirically verified methane emissions of O&G industry at an unprecedented level of accuracy and granularity.

Additionally, IMEO will work in parallel with the Methane Alert and Response System (MARS). MARS will process global satellite data in near real-time to detect and attribute methane emission hot spots. Then, it will notify operators in order to take mitigation actions. MARS is currently in the test phase and the first public results are expected by the beginning of 2024. DIMEI inventory principles could be integrate in the process of automation MARS system. IMEO initiative are crucial for creating the DIMEI inventory, which will require access to the maximum amount of data concerning methane emissions based on a unified data collecting and processing system.

5.2.2 DIMEI for dynamic and intelligent regulations

The monitoring over time of the emission profiles at different scales - which allows the characterization of the behavior of households at different scales

- that DIMEI proposes will allow tracking progress towards the goal of methane mitigation and thus be able to adjust the directives to achieve these objectives. Indeed, the dynamic aspect of the DIMEI opens the possibility for having an adaptive regulatory framework that allows the transition from a generalized regulation to a more efficient personalized regulation system.

Currently, most analyses and decisions on methane emissions are based on ex-post analysis. According to the IEA, both at the industrial and policy level, the goal should be to collect enough initial data and then monitor over time sufficiently to characterise and to anticipate the problem sources. The current framework with the lack of data does not allow to build up emission profiles of sites and their infrastructure which are necessary to predict future emissions. By placing ourselves in the context defined by DIMEI, predictions become possible. Indeed, DIMEI, by its intelligent aspect, not only observes past and current emissions but can predict emissions, which would create a margin of action to prevent any potential accident or failure at the edge of methane emission.

These predictions would introduce the notion of pro-activity into the regulatory system which could permit to adapt designed in advance of suitable mitigation policy for potential future context. It should be noted that all forecasts include a certain margin of error, which must be taken into account in every decision-making process. However, a good knowledge of the emission trends would allow obtaining predictions more or less close to the real emission values. The information obtained from the DIMEI should also allow for evaluation of the effects of policies put in place in order to judge their efficiency and also to understand and analyze the reaction time and reactions of the O&G operators.

Faced with the colossal volume of information that can emerge from the DIMEI, the design of new "personalized regulations" by the usual means could become a cumbersome task and risk losing its optimality. It would be wise to think of a system of recommended regulations and optimal policy in the face of the current or future context. This type of system could support the work of synthesizing the various possible regulatory orientations at local, regional, and national levels.

5.3 Conclusion

The DIMEI, at the proof of concept stage, is able to automatically create methane emission profiles for 3 types of O&G infrastructures (well, tank, compressor) in the Permian Basin (USA). DIMEI is the first AI based inventory regrouping object detection and automatic association, the two mains technical pillars of this thesis. With its snap-shot instantiation module, DIMEI presents the advantage to don't be dependent of O&G infrastructure survey which quickly become inaccurate. Indeed, by regularly using updated satellite images, DIMEI framework could provide up to date infrastructures survey for each detected emission. Launched over time, the DIMEI has the availability to generate emissions profiles at O&G infrastructure level, which could be aggregated from O&G infrastructure to country level (regrouping multiple O&G basins). These emissions profiles are spatio-temporal time series of emitted methane quantities which are keys to characterize emission behavior by different levels. These emissions profiles could even permit to characterize specific action like intentional and non-intentional leak and also be the base of spatio-temporal forecasting of methane emissions. The possibility to forecast methane emissions would permits to create a action margin to avoid these last ones. This forecasting aspect would directly contribute to methane emissions mitigation. DIMEI framework at his stage of proof concept will however need few adjustments to be deploy. It will also required the use of new AI methods to extend his application to other basin than the Permian.

Through DIMEI deployment The DIMEI framework, as showcased in Figure 5.1, necessitates certain refinements to enable its seamless and fully automated operation over time. Specifically, the current iteration of the DIMEI framework mandates manual retrieval of satellite imagery at the precise locations of detected methane emissions for each plume. Incorporating an API to facilitate the automatic extraction of these satellite images is therefore essential for enhancing operational efficiency.

Moreover, to ensure the accuracy of the associations made by DIMEI, it is crucial to employ point source satellite data characterized by a low uncertainty margin regarding the locations of plumes. To comprehensively capture emissions from infrastructures, the satellite data utilized should ideally offer a very low revisit period—permitting multiple scans per day—or, in the optimal scenario, near real-time scanning capabilities. This level of data precision and temporal resolution is vital for DIMEI to function effectively

and deliver reliable results.

Adaptation and extension The DIMEI framework is predicated on employing an object detection algorithm that has been fine-tuned using our OG database. This database comprises satellite images of three types of infrastructures located within the Permian Basin. Consequently, the object detection algorithm, as currently implemented within this framework, is capable of identifying only tanks, wells, and compressors in the Permian Basin (USA).

To broaden the applicability of DIMEI to encompass additional types of infrastructure, such as flares, and to facilitate the detection and recognition of infrastructures in other oil basins, it is imperative to undertake modifications and enhancements to the existing supervised object detection models. This expansion requires an adaptive approach to model training, allowing DIMEI to accurately recognize and analyze a wider array of infrastructure types across diverse geographical regions.

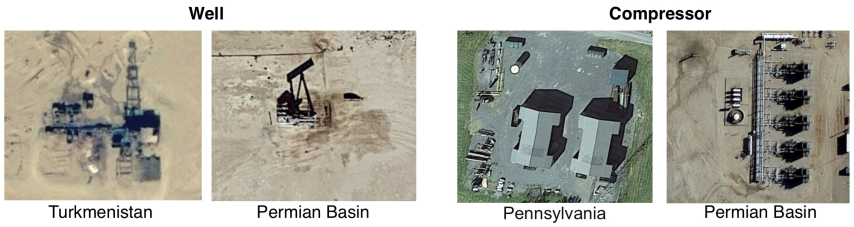


Figure 5.5: Differences between O&G infrastructures by country and by State. *Source : @Google earth.*

Indeed, as depicted in Figure 5.5, the representation of Oil & Gas (O&G) infrastructures can vary significantly from one country to another, and even from one basin to another within the same country. Relying on supervised object detection algorithms necessitates their recalibration for each new set of infrastructures and for each new locale characterized by distinct infrastructure types. This calibration process demands the creation of an extensive manually annotated image database, which is both time-consuming and suboptimal. To circumvent the need for this labor-intensive supervised fine-tuning phase, exploring the potential of few-shot and self-supervised learning algorithms for object detection emerges as a viable alternative. These innovative approaches, capable of detecting and recognizing new and

varied infrastructures, require few or no annotated images, thereby facilitating the seamless expansion of the DIMEI project to encompass new infrastructures and geographical areas with greater efficiency.

Chapter 6

Conclusion

In this thesis, we introduce a novel methane emission inventory utilizing artificial intelligence. The Dynamic and Intelligent Methane Emissions Inventory (DIMEI) facilitates automated monitoring of methane emissions from three types of oil and gas (O&G) infrastructure in the Permian basin (USA). As the DIMEI accumulates data over time, it captures methane emission quantities from wells, tanks, and compressors, reflecting the emissions profile of these infrastructures (dynamic component). This longitudinal data collection enables characterization of the emission behavior across various levels, including site, basin, country, and operator, and supports forecasting of methane emissions (intelligent component).

The forecasting capability provided by the intelligent component creates a proactive window for intervention to prevent methane emissions, thereby directly contributing to methane mitigation efforts. Additionally, the emissions profile information, ranging from infrastructure to operator level, could inform the direction and design of methane mitigation regulations.

The DIMEI inventory is based on two principal pillars:

1. Automated O&G infrastructures detection and recognition in the Permian basin (USA) (Chapter 3);

2. Automated association of detected methane plumes to O&G infrastructure, sites and operators at their origin (Chapter 4).

6.1 Contributions

We meticulously fine-tuned and assessed three categories of object detection algorithms specifically designed for the automated identification and monitoring of three types of Oil & Gas (O&G) infrastructures within the Permian Basin, USA. To achieve this, we developed the inaugural database of O&G infrastructure images by meticulously annotating high-resolution satellite imagery.

Our research includes a comprehensive quantitative analysis of the impact of pre-training on the detection capabilities of these algorithms, comparing their performance with and without pre-training. Furthermore, we delved into the resilience of these three algorithms by systematically quantifying the effects of common satellite image disturbances on their accuracy and performance.

In Chapter 4, we introduced a novel framework developed to facilitate the automatic attribution of detected methane plumes to specific types of Oil & Gas (O&G) sites and their operators. Utilizing a clustering algorithm, this framework successfully automated the linkage of over 100 detected methane plumes, identified via point source satellite imaging, to various O&G site types and operators across the Permian Basin. Indirectly, the framework also contributed to the temporal extension of the PermianMAP survey by enabling the automatic association of new methane emissions with previously surveyed sites. We concluded the chapter by presenting a method for the automated attribution of detected methane plumes to O&G infrastructures and discussed potential approaches to manage the uncertainties associated with these attributions.

In Chapter 5, we have presented the first methane emission inventory "Dynamic and Intelligent Methane Inventory" (DIMEI) based on AI which combines the methods presented in Chapter 3 and Chapter 4. We have demonstrated his ability to automatically follow in time the methane emissions from Permian O&G infrastructures (dynamic component). We have also presented and detailed the possible future automated analysis (intelligent component) that could be conduct based on use of DIMEI mid-log term outputs. We concluded theses works by discussing the symbiosis between the DIMEI and regulations for optimized methane mitigation.

6.2 Future work

The research surrounding DIMEI suggests several avenues for short-term improvement and long-term development.

6.2.1 Short-term Perspective

Object Detection: Performance and Robustness The object detection algorithms employed are based on supervised learning, necessitating the creation of a dedicated database for the fine-tuning phase of the three algorithms.

Given that the extraction and annotation of O&G site satellite images are time-consuming tasks, our database currently comprises 930 images and 1951 annotations, which is relatively small for the domain. Expanding our database could enhance the detection and recognition capabilities of the three algorithms. This can be achieved by extracting and annotating new images or employing remote sensing data augmentation methods. These methods involve using existing images to generate modified images through rotations, blurring, noise addition, etc. The parameters for generating new images can be determined based on our findings regarding the impact of satellite image noise on algorithm performance. Specifically, generating images with the most impactful noise could bolster algorithm robustness through fine-tuning on this new dataset.

Our experiments have indicated that each of the three object detection algorithms has its own advantages and limitations. To capitalize on the strengths of these algorithms and mitigate their shortcomings, ensemble models could be employed to combine them.

DIMEI Extension and Adaptability While our current work concentrates on determining the emission profiles of tanks, compressors, and wells in the Permian Basin, it is essential to consider other types of infrastructure, such as flares, which are significant methane emission sources [84]. Additionally, differences in infrastructure appearance across different O&G basins may necessitate analysis beyond the Permian Basin. To facilitate the extension of DIMEI to other infrastructures without the need to manually construct a new database, few-shot or self-supervised learning methods

could be beneficial. Our ongoing research aims to leverage these methods to expedite the extension of DIMEI to diverse infrastructures.

Association: Uncertainty Management Our research on associating detected methane plumes with O&G infrastructures and sites currently lacks the incorporation of uncertainties related to the detected plume positions. The database of source point detections we use does not include detailed information regarding the uncertainties associated with either the estimated quantity of emitted methane or the estimated location. Incorporating the percentage of uncertainty in the estimated location for each plume would enable a more rigorous attribution process, leveraging probabilities for a plume's association with other infrastructures.

6.2.2 Long-term Perspective

O&G Sites Types Automated Detection The framework we propose for automatically associating methane plumes with O&G site types and operators relies on existing surveys, such as the PermianMAP project in our case. This framework necessitates a database containing site locations, types, and operators. While databases like OGIM and GOGi provide such information, they quickly become outdated due to the rapidly evolving O&G landscape. Although methods like METER-ML and OGNet can detect oil refineries and petroleum terminals, further refinement is required to precisely identify site types, such as production, gathering, boosting, and processing. Our dynamic snapshot method for infrastructures could potentially be adapted to identify the type of O&G site without depending on existing static surveys.

O&G Operators Site Automated Survey Unlike determining the type of O&G site, identifying the operator of an O&G site cannot be achieved through object detection methods due to the absence of distinct visual indicators. While several databases offer information about site operators, they are not freely accessible. A potential solution to obtain operator information could involve automatically extracting data from the Title V program [117] website through web scraping. This program provides access to permit applications for each O&G site infrastructure, including their addresses. However, this approach has limitations as a permit application does not always correspond to the actual location of the infrastructure.

DIMEI Future Deployment and Use The DIMEI framework, currently at the proof-of-concept stage, requires several adjustments to be deployed and to provide continuous, long-term methane emissions profiles. Our framework utilizes a single satellite data source for feature detections, but integrating various satellite sources with different estimates could enhance detection capabilities. Utilizing an API to automatically and efficiently extract satellite images corresponding to each methane detection location could also be beneficial.

In the medium to long term, the intelligent aspect of DIMEI could be leveraged to generate substantial amounts of data, i.e., emission profiles. Given the large volumes of data this new inventory type will produce, manually transforming this data into actionable information for methane mitigation regulation design will be challenging for policymakers. This shift in the inventory landscape necessitates an evolution in policy design frameworks, which will need to incorporate artificial intelligence to efficiently process this wealth of new information. Consequently, new regulatory frameworks tailored to the use of innovative inventory types like DIMEI will be essential.

Bibliography

[1] Ramón A. Alvarez et al. "Assessment of methane emissions from the U.S. oil and gas supply chain". In: *Science* 361.6398 (2018), pp. 186–188. DOI: 10.1126/science.aar7204. eprint: https://www.science.org/doi/pdf/10.1126/science.aar7204. URL: https://www.science.org/doi/abs/10.1126/science.aar7204.

[2] Franz Aurenhammer and Rolf Klein. "Chapter 5 - Voronoi Diagrams**Partially supported by the Deutsche Forschungsgemeinschaft, grant K1 655 2-2." In: *Handbook of Computational Geometry*. Ed. by J.-R. Sack and J. Urrutia. Amsterdam: North-Holland, 2000, pp. 201–290. ISBN: 978-0-444-82537-7. DOI: https://doi.org/10.1016/B978-044482537-7/50006-1. URL: https://www.sciencedirect.com/science/article/pii/B9780444825377500061.

[3] Justin Bagley et al. "Assessment of an atmospheric transport model for annual inverse estimates of California greenhouse gas emissions". In: *Journal of Geophysical Research Atmospheres* 122 (Feb. 2017). DOI: 10.1002/2016JD025361.

[4] Pascal Barthe et al. "Best available techniques (BAT) reference document for the refining of mineral oil and gas". In: *European Commission* 754 (2015).

[5] David Beerling et al. "Methane and the CH 4 related greenhouse effect over the past 400 million years". In: *American Journal of Science* 309.2 (2009), pp. 97–113.

[6] D. A. Belikov et al. "Adjoint of the global Eulerian–Lagrangian coupled atmospheric transport model (A-GELCA v1.0): development and validation". In: *Geoscientific Model Development* 9.2 (2016), pp. 749–764. DOI: 10.5194/gmd-9-749-2016. URL: https://gmd.copernicus.org/articles/9/749/2016/.

[7] Kanchan Bhil et al. "Recent Progress in Object Detection in Satellite Imagery: A Review". In: *Sustainable Advanced Computing*. Ed. by Sagaya Aurelia et al. Singapore: Springer Singapore, 2022, pp. 209–218. ISBN: 978-981-16-9012-9.

[8] J. Bruno et al. "U-Plume: Automated algorithm for plume detection and source quantification by satellite point-source imagers". In: *EGUsphere* 2023 (2023), pp. 1–24. DOI: 10.5194/egusphere-2023-1343. URL: https://egusphere.copernicus.org/preprints/2023/egusphere-2023-1343/.

[9] M. Buchwitz et al. "Satellite-derived methane hotspot emission estimates using a fast data-driven method". In: *Atmospheric Chemistry and Physics* 17.9 (2017), pp. 5751–5774. DOI: 10.5194/acp-17-5751-2017. URL: https://acp.copernicus.org/articles/17/5751/2017/.

[10] Nicolas Carion et al. "End-to-end object detection with transformers". In: (2020), pp. 213–229.

[11] Elton Chan et al. "Eight-Year Estimates of Methane Emissions from Oil and Gas Operations in Western Canada Are Nearly Twice Those Reported in Inventories". In: *Environmental Science & Technology* 54.23 (2020). PMID: 33169990, pp. 14899–14909. DOI: 10.1021/acs.est.0c04117. eprint: https://doi.org/10.1021/acs.est.0c04117. URL: https://doi.org/10.1021/acs.est.0c04117.

[12] S. Conley et al. "Application of Gauss's theorem to quantify localized surface emissions from airborne measurements of wind and trace gases". In: *Atmospheric Measurement Techniques* 10.9 (2017), pp. 3345–3358. DOI: 10.5194/amt-10-3345-2017. URL: https://amt.copernicus.org/articles/10/3345/2017/.

[13] U Cubasch et al. "G. Introduction, in". In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group 1 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (), pp. 119–158.

[14] Daniel H. Cusworth et al. "Intermittency of Large Methane Emitters in the Permian Basin". In: *Environmental Science & Technology Letters* 8.7 (2021), pp. 567–573. DOI: 10.1021/acs.estlett.1c00173. eprint: https://doi.org/10.1021/acs.estlett.1c00173. URL: https://doi.org/10.1021/acs.estlett.1c00173.

[15] Daniel H. Cusworth et al. "Strong methane point sources contribute a disproportionate fraction of total emissions across multiple basins in the United States". In: *Proceedings of the National Academy of Sciences* 119.38 (2022), e2202338119. DOI: 10.1073/pnas.2202338119. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2202338119. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2202338119.

[16] Navneet Dalal and Bill Triggs. "Histograms of oriented gradients for human detection". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* 1 (2005), 886–893 vol. 1.

[17] Pratyush Datta. "How effective LDAR campaigns contribute to minimizing methane emissions". In: *Abu Dhabi International Petroleum Exhibition and Conference*. SPE. 2020, D021S029R004.

[18] Dmitry Demidov, Rushali Grandhe, and Salem Almarri. "Object Detection in Aerial Imagery". In: (Nov. 2022). DOI: 10.48550/arXiv.2211.15479.

[19] Jia Deng et al. "Imagenet: A large-scale hierarchical image database". In: Ieee. 2009, pp. 248–255.

[20] Z. Deng et al. "Comparing national greenhouse gas budgets reported in UNFCCC inventories against atmospheric inversions". In: *Earth System Science Data* 14.4 (2022), pp. 1639–1675. DOI: 10.5194/essd-14-1639-2022. URL: https://essd.copernicus.org/articles/14/1639/2022/.

[21] Philip M Dixon. "Nearest neighbor methods". In: *Encyclopedia of environmetrics* 3 (2002), pp. 1370–1383.

[22] Edward J Dlugokencky et al. "Global atmospheric methane: budget, changes and dangers". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369.1943 (2011), pp. 2058–2072.

[23] Débora F. Dos Santos et al. "Performance Comparison of Convolutional Neural Network Models for Object Detection in Tethered Balloon Imagery". In: *2021 Latin American Robotics Symposium (LARS), 2021 Brazilian Symposium on Robotics (SBR), and 2021 Workshop on Robotics in Education (WRE)*. 2021, pp. 246–251. DOI: 10.1109/LARS/SBR/WRE54079.2021.9605459.

[24] Guillaume Druart et al. "Test of SIMAGAZ: a LWIR cryogenic multispectral infrared camera for methane gas leak detection and quantification". In: *Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXVII*. Vol. 11727. SPIE. 2021, pp. 53–59.

[25] Martin Ester et al. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. In: KDD’96. Portland, Oregon: AAAI Press, 1996, 226–231.

[26] Mark Everingham et al. “The Pascal Visual Object Classes (VOC) Challenge”. In: *Int. J. Comput. Vision* 88.2 (June 2010), 303–338. ISSN: 0920-5691. DOI: 10.1007/s11263-009-0275-4. URL: https://doi.org/10.1007/s11263-009-0275-4.

[27] Forrest Fankhauser, J. Anthony Tyson, and Jacob Askari. “Satellite Optical Brightness”. In: *The Astronomical Journal* 166.2 (July 2023), p. 59. DOI: 10.3847/1538-3881/ace047. URL: https://dx.doi.org/10.3847/1538-3881/ace047.

[28] Boris Gašparović et al. “Deep Learning Approach For Objects Detection in Underwater Pipeline Images”. In: *Applied Artificial Intelligence* 36.1 (2022), p. 2146853. DOI: 10.1080/08839514.2022.2146853. eprint: https://doi.org/10.1080/08839514.2022.2146853. URL: https://doi.org/10.1080/08839514.2022.2146853.

[29] C. Geels et al. “Comparing atmospheric transport models for future regional inversions over Europe ; Part 1: mapping the atmospheric CO2 signals”. In: *Atmospheric Chemistry and Physics* 7.13 (2007), pp. 3461–3479. DOI: 10.5194/acp-7-3461-2007. URL: https://acp.copernicus.org/articles/7/3461/2007/.

[30] Ross Girshick. “Fast r-cnn”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1440–1448.

[31] Ross Girshick et al. “Rich feature hierarchies for accurate object detection and semantic segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 580–587.

[32] GlobalMethaneTracker. *https://iea.blob.core.windows.net/assets/b5f6bb13-76ce-48ea-8fdb-3d4f8b58c838/GlobalMethaneTracker*$_d$*ocumentation.pdf*.

[33] Joost de Gouw et al. “Daily Satellite Observations of Methane from Oil and Gas Production Regions in the United States”. In: *Scientific Reports* 10 (Jan. 2020). DOI: 10.1038/s41598-020-57678-4.

[34] Jade E Guisiano et al. “O&GProfile : An automated method for attribution of satellite methane emissions detections to oil and gas sites and operators”. In: *International Conference on Machine Learning and Data Mining MLDM 2023*. New-York, United States, July 2023. URL: https://hal.science/hal-04043407.

[35] Jade Eva Guisiano et al. "Dynamic & Intelligent Methane Emissions Inventory (DIMEI) Framework: Next-generation methane emission inventory for oil and gas industry based on Artificial Intelligence". In: *iScience* (2024).

[36] Jade Eva Guisiano et al. "Object detection models sensitivity & robustness
to satellite-based adversarial attacks". In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Athens, Greece, 2024.

[37] Jade Eva Guisiano et al. "Oil and Gas Automatic Infrastructure Mapping: Leveraging High-Resolution Satellite Imagery through fine-tuning of object detection models". In: *International Conference on Neural Information Processing (ICONIP)*. Changsha, China, 2023. URL: https://hal.science/hal-04197007.

[38] Kemal Hacıefendioğlu, Hasan Basri Başağa, and Gökhan Demir. "Automatic detection of earthquake-induced ground failure effects through Faster R-CNN deep learning-based object detection using satellite images". In: *Natural Hazards* 105 (2021), pp. 383–403.

[39] J. Hansen, M. Sato, and R. Ruedy. "Radiative forcing and climate response". In: *J. Geophys. Res.* 102 (1997), pp. 6831–6864. DOI: 10.1029/96JD03436.

[40] Gabriel Huang et al. "A survey of self-supervised and few-shot object detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

[41] IEA. "IEA Methane Tracker". In: (2023). URL: https://www.iea.org/data-and-statistics/data-tools/methane-tracker.

[42] IPCC. "Climate Change 2013: The Physical Science Basis (AR5)". In: (2013). URL: https://www.ipcc.ch/report/ar5/wg1/.

[43] IPCC. "IPCC AR6 Report". In: (2023). URL: https://www.ipcc.ch/report/ar6/wg1/figures/summary-for-policymakers/figure-spm-2.

[44] Itziar Irakulis-Loitxate et al. "Satellite-based survey of extreme methane emissions in the Permian basin". In: *Science Advances* 7.27 (2021), eabf4507. DOI: 10.1126/sciadv.abf4507. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.abf4507. URL: https://www.science.org/doi/abs/10.1126/sciadv.abf4507.

[45] D. J. Jacob et al. "Quantifying methane emissions from the global scale down to point sources using satellite observations of atmospheric methane". In: *Atmospheric Chemistry and Physics Discussions* 2022 (2022), pp. 1–44. DOI: 10.5194/acp-2022-246. URL: https://acp.copernicus.org/preprints/acp-2022-246/.

[46] Maroš Jakubec et al. "Comparison of CNN-Based Models for Pothole Detection in Real-World Adverse Conditions: Overview and Evaluation". In: *Applied Sciences* 13.9 (2023). ISSN: 2076-3417. DOI: 10.3390/app13095810. URL: https://www.mdpi.com/2076-3417/13/9/5810.

[47] Xin Jin and Jiawei Han. "K-Means Clustering". In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, 2010, pp. 563–564. ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_425. URL: https://doi.org/10.1007/978-0-387-30164-8_425.

[48] Derek Johnson et al. "Methane emissions from oil and gas production sites and their storage tanks in West Virginia". In: *Atmospheric Environment: X* 16 (2022), p. 100193. ISSN: 2590-1621. DOI: https://doi.org/10.1016/j.aeaoa.2022.100193. URL: https://www.sciencedirect.com/science/article/pii/S2590162122000478.

[49] Matthew Johnson, David Tyner, and Bradley Conrad. "Origins of Oil and Gas Sector Methane Emissions: On-Site Investigations of Aerial Measured Sources". In: *Environmental Science & Technology* 57 (Jan. 2023). DOI: 10.1021/acs.est.2c07318.

[50] Siraput Jongaramrungruang et al. "MethaNet – An AI-driven approach to quantifying methane point-source emission from high-resolution 2-D plume imagery". In: *Remote Sensing of Environment* 269 (2022), p. 112809. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2021.112809. URL: https://www.sciencedirect.com/science/article/pii/S0034425721005290.

[51] Junhyung Kang et al. "A Survey of Deep Learning-Based Object Detection Methods and Datasets for Overhead Imagery". In: *IEEE Access* 10 (2022), pp. 20118–20134. DOI: 10.1109/ACCESS.2022.3149052.

[52] Seyedahmad Kia et al. "Machine Learning to Predict Area Fugitive Emission Fluxes of GHGs from Open-Pit Mines". In: *Atmosphere* 13.2 (2022). ISSN: 2073-4433. URL: https://www.mdpi.com/2073-4433/13/2/210.

[53] Juhyun Kim et al. "The development of leak detection model in subsea gas pipeline using machine learning". In: *Journal of Natural Gas Science and Engineering* 94 (2021), p. 104134. ISSN: 1875-5100. DOI: https://doi.org/10.1016/j.jngse.2021.104134.

[54] Alibek Kopbayev et al. "Gas leakage detection using spatial and temporal neural network model". In: *Process Safety and Environmental Protection* 160 (2022), pp. 968–975. ISSN: 0957-5820. DOI: https://doi.org/10.1016/j.psep.2022.03.002.

[55] Y. Koyama et al. "Simulation of variability in atmospheric carbon dioxide using a global coupled Eulerian â€" Lagrangian transport model". In: *Geoscientific Model Development* 4.2 (2011), pp. 317–324. DOI: 10.5194/gmd-4-317-2011. URL: https://gmd.copernicus.org/articles/4/317/2011/.

[56] G. Kuhlmann et al. "Quantifying CO2 emissions of a city with the Copernicus Anthropogenic CO2 Monitoring satellite mission". In: *Atmospheric Measurement Techniques* 13.12 (2020), pp. 6733–6754. DOI: 10.5194/amt-13-6733-2020. URL: https://amt.copernicus.org/articles/13/6733/2020/.

[57] Guanter L. et al. "Methane Alert and response system (MARS): IMEO's satellite-based system for detection and attribution of methane point sources around the world". In: *EGU General Assembly 2023* (2022). DOI: https://doi.org/10.5194/egusphere-egu23-9548.

[58] T. Lauvaux et al. "Global assessment of oil and gas methane ultra-emitters". In: *Science* 375.6580 (Feb. 2022), 557–561. ISSN: 1095-9203. DOI: 10.1126/science.abj4351. URL: http://dx.doi.org/10.1126/science.abj4351.

[59] Mohamed Layouni, Mohamed Salah Hamdi, and Sofiène Tahar. "Detection and sizing of metal-loss defects in oil and gas pipelines using pattern-adapted wavelets and machine learning". In: *Applied Soft Computing* 52 (2017), pp. 247–261. ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2016.10.040. URL: https://www.sciencedirect.com/science/article/pii/S1568494616305658.

[60] J. Lelieveld, P.J. Crutzen, and C. Brühl. "Climate effects of atmospheric methane". In: *Chemosphere* 26.1 (1993). Proceedings of the NATO advanced research workshop, pp. 739–768. ISSN: 0045-6535. DOI: https://doi.org/10.1016/0045-6535(93)90458-H. URL: https://www.sciencedirect.com/science/article/pii/004565359390458H.

[61] Jos Lelieveld and Paul J Crutzen. "Indirect chemical effects of methane on climate warming". In: *Nature* 355.6358 (1992), pp. 339–342.

[62] Debang Li, Junge Zhang, and Kaiqi Huang. "Universal adversarial perturbations against object detection". In: *Pattern Recognition* 110 (2021), p. 107584. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2020.107584. URL: https://www.sciencedirect.com/science/article/pii/S0031320320303873.

[63] Min Li et al. "Agricultural Greenhouses Detection in High-Resolution Satellite Images Based on Convolutional Neural Networks: Comparison of Faster R-CNN, YOLO v3 and SSD". In: *Sensors* 20.17 (2020). ISSN: 1424-8220. DOI: 10.3390/s20174938. URL: https://www.mdpi.com/1424-8220/20/17/4938.

[64] Qingyun Li, Yushi Chen, and Ying Zeng. "Transformer with Transfer CNN for Remote-Sensing-Image Object Detection". In: *Remote Sensing* 14.4 (2022). ISSN: 2072-4292. DOI: 10.3390/rs14040984. URL: https://www.mdpi.com/2072-4292/14/4/984.

[65] Zheng Li et al. "Deep Learning-Based Object Detection Techniques for Remote Sensing Images: A Survey". In: *Remote Sensing* 14.10 (2022). ISSN: 2072-4292. DOI: 10.3390/rs14102385. URL: https://www.mdpi.com/2072-4292/14/10/2385.

[66] Tsung-Yi Lin et al. "Focal loss for dense object detection". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2980–2988.

[67] Tsung-Yi Lin et al. "Microsoft COCO: Common Objects in Context". In: (2014). Ed. by David Fleet et al., pp. 740–755.

[68] Tony Lindeberg. "Scale Invariant Feature Transform". In: vol. 7. May 2012. DOI: 10.4249/scholarpedia.10491.

[69] Wei Liu et al. "Ssd: Single shot multibox detector". In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer. 2016, pp. 21–37.

[70] Hongfang Lu et al. "Oil and Gas 4.0 era: A systematic review and outlook". In: *Computers in Industry* 111 (2019), pp. 68–90.

[71] J. D. Maasakkers et al. "2010–2015 North American methane emissions, sectoral contributions, and trends: a high-resolution inversion of GOSAT observations of atmospheric methane". In: *Atmospheric Chemistry and Physics* 21.6 (2021), pp. 4339–4356. DOI: 10.5194/acp-21-4339-2021. URL: https://acp.copernicus.org/articles/21/4339/2021/.

[72] J. D. Maasakkers et al. "Global distribution of methane emissions, emission trends, and OH concentrations and trends inferred from an inversion of GOSAT satellite data for 2010–2015". In: *Atmospheric Chemistry and Physics* 19.11 (2019), pp. 7859–7881. DOI: 10.5194/acp-19-7859-2019. URL: https://acp.copernicus.org/articles/19/7859/2019/.

[73] Kathleen Mar et al. "Beyond CO2 equivalence: The impacts of methane on climate, ecosystems, and health". In: *Environmental Science & Policy* 134 (Aug. 2022), pp. 127–136. DOI: 10.1016/j.envsci.2022.03.027.

[74] Mahmoud Meribout et al. "Leak detection systems in oil and gas fields: Present trends and future prospects". In: *Flow Measurement and Instrumentation* 75 (2020), p. 101772.

[75] Mehdi Mohammadpoor and Farshid Torabi. "Big Data analytics in oil and gas industry: An emerging trend". In: *Petroleum* 6.4 (2020), pp. 321–328.

[76] MBME Muhammad, IR Ehigiator, and SO Oladosu. "Application of LIDAR Technology in Oil and Gas Pipeline Route Selection and Optimization". In: *J Remote Sens GIS* 10.5 (2021), p. 288.

[77] NationalAcademiesofSciencesEngineeringandMedicine. "Improving Characterization of Anthropogenic Methane Emissions in the United States". In: *The National Academies Pres* (2018). DOI: https://doi.org/10.17226/24987.

[78] Stijn Naus et al. "Assessing the Relative Importance of Satellite-Detected Methane Superemitters in Quantifying Total Emissions for Oil and Gas Production Areas in Algeria". In: *Environmental Science & Technology* 57 (Nov. 2023). DOI: 10.1021/acs.est.3c04746.

[79] NOAA. "NOAA's 2023 Annual Climate Report". In: (2023). URL: http://elib.suub.uni-bremen.de/publications/dissertations/E-Diss845_treude.pdf.

[80] Sudhanshu Pandey et al. "Automated monitoring of methane super-emitters using multispectral satellite instruments and machine learning." In: *AGU Fall Meeting Abstracts*. Vol. 2021. Dec. 2021, A54F-02, A54F–02.

[81] D. Picard et al. *FUGITIVE EMISSIONS*. The Intergovernmental Panel on Climate Change, 2006. Chap. 4.

[82] D. Pillai et al. "Comparing Lagrangian and Eulerian models for CO2 transport â€“ a step towards Bayesian inverse modeling using WRF/STILT-VPRM". In: *Atmospheric Chemistry and Physics* 12.19 (2012), pp. 8979–8991. DOI: 10.5194/acp-12-8979-2012. URL: https://acp.copernicus.org/articles/12/8979/2012/.

[83] Ignacio Pisso et al. "Assessing Lagrangian inverse modelling of urban anthropogenic CO2 fluxes using in situ aircraft and ground-based measurements in the Tokyo area". In: *Carbon Balance and Management* 14 (May 2019). DOI: 10.1186/s13021-019-0118-8.

[84] Genevieve Plant et al. "Inefficient and unlit natural gas flares both emit large quantities of methane". In: *Science* 377.6614 (2022), pp. 1566–1571. DOI: 10.1126/science.abq0385. eprint: https://www.science.org/doi/pdf/10.1126/science.abq0385. URL: https://www.science.org/doi/abs/10.1126/science.abq0385.

[85] Prajaya Prajapati and Eduardo A Santos. "Comparing methane emissions estimated using a backward-Lagrangian stochastic model and the eddy covariance technique in a beef cattle feedlot". In: *Agricultural and forest meteorology* 256 (2018), pp. 482–491.

[86] Nilantha Premakumara et al. "Improving Object Detection Robustness against Natural Perturbations through Synthetic Data Augmentation". In: *Proceedings of the 2023 Asia Conference on Computer Vision, Image Processing and Pattern Recognition*. CVIPPR '23. Phuket,Thailand: Association for Computing Machinery, 2023. ISBN: 9798400700033. DOI: 10.1145/3596286.3596293.

[87] Ali Radman et al. "S2MetNet: A novel dataset and deep learning benchmark for methane point source quantification using Sentinel-2 satellite imagery". In: *Remote Sensing of Environment* 295 (2023), p. 113708. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2023.113708. URL: https://www.sciencedirect.com/science/article/pii/S0034425723002596.

[88] S. Rajkumar and Malathi Ganesan. "A Comparative Analysis on Image Quality Assessment for Real Time Satellite Images". In: *Indian Journal of Science and Technology* 9 (Sept. 2016). DOI: 10.17485/ijst/2016/v9i34/96766.

[89] Subramanian Ramachandran et al. "Methane Emissions from Natural Gas Compressor Stations in the Transmission and Storage Sector: Measurements and Comparisons with the EPA Greenhouse Gas Reporting Program Protocol". In: *Environmental science & technology* 49 (Feb. 2015). DOI: 10.1021/es5060258.

[90] Sebastian Ramiro-Ramirez et al. "Porosity and Permeability Heterogeneity in the Upper Wolfcamp, Delaware Basin, West Texas: Implications for Production". In: July 2020. DOI: 10.15530/urtec-2020-2105.

[91] Joseph Redmon et al. "You only look once: Unified, real-time object detection". In: (2016), pp. 779–788.

[92] Shaoqing Ren et al. "Faster r-cnn: Towards real-time object detection with region proposal networks". In: *Advances in neural information processing systems* 28 (2015).

[93] Stuart N. Riddick and Denise L. Mauzerall. "Likely substantial underestimation of reported methane emissions from United Kingdom upstream oil and gas activities". In: *Energy Environ. Sci.* 16 (1 2023), pp. 295–304. DOI: 10.1039/D2EE03072A. URL: http://dx.doi.org/10.1039/D2EE03072A.

[94] M. Rigby, A. J. Manning, and R. G. Prinn. "Inversion of long-lived trace gas emissions using combined Eulerian and Lagrangian chemical transport models". In: *Atmospheric Chemistry and Physics* 11.18 (2011), pp. 9887–9898. DOI: 10.5194/acp-11-9887-2011. URL: https://acp.copernicus.org/articles/11/9887/2011/.

[95] Gautam Ritesh. "Oil and Gas Infrastructure Mapping (OGIM) database". Version OGIM_v1.1. In: (May 2023). DOI: 10.5281/zenodo.7922117. URL: https://doi.org/10.5281/zenodo.7922117.

[96] Clive D. Rodgers. *Inverse methods for atmospheric sounding : theory and practice*. Ed. by Clive D. Rodgers. World Scientific Publishing, 2000.

[97] Jeffrey Rutherford et al. "Closing the methane gap in US oil and natural gas production emissions inventories". In: *Nature Communications* 12 (Aug. 2021). DOI: 10.1038/s41467-021-25017-4.

[98] Michael Sabbatino. "Global Oil & Gas Infrastructure Features Database Geocube Collection". In: (Mar. 2018). DOI: 10.18141/1502839. URL: https://www.osti.gov/biblio/1502839.

[99] Pankaj Sadavarte et al. "Methane Emissions from Superemitting Coal Mines in Australia Quantified Using TROPOMI Satellite Observations". In: *Environmental Science & Technology* 55.24 (2021). PMID: 34842427, pp. 16573–16580. DOI: 10.1021/acs.est.1c03976. eprint: https://doi.org/10.1021/acs.est.1c03976. URL: https://doi.org/10.1021/acs.est.1c03976.

[100] M. Saunois et al. "The Global Methane Budget 2000–2017". In: *Earth System Science Data* 12.3 (2020), pp. 1561–1623. DOI: 10.5194/essd-12-1561-2020. URL: https://essd.copernicus.org/articles/12/1561/2020/.

[101] B. J. Schuit et al. "Automated detection and monitoring of methane super-emitters using satellite data". In: *Atmospheric Chemistry and Physics Discussions* 2023 (2023), pp. 1–47. DOI: 10.5194/acp-2022-862. URL: https://acp.copernicus.org/preprints/acp-2022-862/.

[102] A. Shah et al. "Testing the near-field Gaussian plume inversion flux quantification technique using unmanned aerial vehicle sampling". In: *Atmospheric Measurement Techniques* 13.3 (2020), pp. 1467–1484. DOI: 10.5194/amt-13-1467-2020. URL: https://amt.copernicus.org/articles/13/1467/2020/.

[103] Adil Shah et al. "A Near-Field Gaussian Plume Inversion Flux Quantification Method, Applied to Unmanned Aerial Vehicle Sampling". In: *Atmosphere* 10.7 (2019). ISSN: 2073-4433. DOI: 10.3390/atmos10070396. URL: https://www.mdpi.com/2073-4433/10/7/396.

[104] Lu Shen et al. "Unravelling a large methane emission discrepancy in Mexico using satellite observations". In: *Remote Sensing of Environment* 260 (2021), p. 112461. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2021.112461. URL: https://www.sciencedirect.com/science/article/pii/S0034425721001796.

[105] Hao Sheng et al. "OGNet: Towards a Global Oil and Gas Infrastructure Database using Deep Learning on Remotely Sensed Imagery". In: *ArXiv* abs/2011.07227 (2020).

[106] Jihao Shi et al. "Real-time natural gas release forecasting by using physics-guided deep learning probability model". In: *Journal of Cleaner Production* 368 (2022), p. 133201. ISSN: 0959-6526. DOI: https://doi.org/10.1016/j.jclepro.2022.133201. URL: https://www.sciencedirect.com/science/article/pii/S0959652622027895.

[107] Pengfei Shi et al. "Oil Well Detection via Large-Scale and High-Resolution Remote Sensing Images Based on Improved YOLO v4". In: *Remote Sensing* 13.16 (2021). ISSN: 2072-4292. DOI: 10.3390/rs13163243. URL: https://www.mdpi.com/2072-4292/13/16/3243.

[108] Guanfu Song et al. "Detection of oil wells based on faster R-CNN in optical satellite remote sensing images". In: *Image and Signal Processing for Remote Sensing XXVI*. Vol. 11533. SPIE. 2020, pp. 114–121.

[109] Sarvesh Kumar Sonkar et al. "Detection and Estimation of Natural Gas Leakage Using UAV by Machine Learning Algorithms". In: *IEEE Sensors Journal* 22.8 (2022), pp. 8041–8049. DOI: 10.1109/JSEN.2022.3157872.

[110] Huiming Sun et al. "Defense against Adversarial Cloud Attack on Remote Sensing Salient Object Detection". In: (2023). arXiv: 2306.17431 [cs.CV].

[111] Arsalan Tahir et al. "Automatic Target Detection from Satellite Imagery Using Machine Learning". In: *Sensors* 22.3 (2022). ISSN: 1424-8220. DOI: 10.3390/s22031147. URL: https://www.mdpi.com/1424-8220/22/3/1147.

[112] Lu Tan et al. "Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification". In: *BMC Medical Informatics and Decision Making* 21 (Nov. 2021). DOI: 10.1186/s12911-021-01691-8.

[113] Juan Terven and Diana-Margarita Cordova-Esparza. "A Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond". In: (Apr. 2023).

[114] TheWorldBank. "World Bank Group Is Leading the Effort on Methane Emissions Reduction with Impactful Projects and Initiativest". In: (2022). URL: https://www.worldbank.org/en/news/factsheet/2022/09/19/world-bank-group-is-leading-the-effort-on-methane-emissions-reduction-with-impactful-projects-and-initiatives.

[115] Ranga Rajan Thiruvenkatachari et al. "Uncertainty in using dispersion models to estimate methane emissions from manure lagoons in dairies". In: *Agricultural and Forest Meteorology* 290 (2020), p. 108011.

[116] Xinghao Tian et al. "Leakage detection of low-pressure gas distribution pipeline system based on linear fitting and extreme learning machine". In: *International Journal of Pressure Vessels and Piping* 194 (2021), p. 104553. ISSN: 0308-0161. DOI: https://doi.org/10.1016/j.ijpvp.2021.104553.

[117] TitleV. "Title V Operating Permits". In: (). URL: https://www.epa.gov/title-v-operating-permits.

[118] Tina Treude. "Anaerobic Oxidation of Methane in marine sediments." In: (Jan. 2004). URL: http://elib.suub.uni-bremen.de/publications/dissertations/E-Diss845_treude.pdf.

[119] Erin E Tullos et al. "Use of short duration measurements to estimate methane emissions at oil and gas production sites". In: *Environmental Science & Technology Letters* 8.6 (2021), pp. 463–467.

[120] Alexander J Turner, Christian Frankenberg, and Eric A Kort. "Interpreting contemporary trends in atmospheric methane". In: *Proceedings of the National Academy of Sciences* 116.8 (2019), pp. 2805–2813.

[121] David Tyner and Matthew Johnson. "Where the Methane Is—Insights from Novel Airborne LiDAR Measurements Combined with Ground Survey Data". In: *Environmental Science and Technology* 55 (July 2021), 9773–9783. DOI: 10.1021/acs.est.1c01572.

[122] UNEP. "Global Methane Assessment: 2030 Baseline Report". In: (2022). URL: https://www.unep.org/resources/report/global-methane-assessment-2030-baseline-report.

[123] USGCRP. "Climate Science Special Report". In: (2017). URL: https://science2017.globalchange.gov/downloads/CSSR2017_FullReport.pdf.

[124] D. J. Varon et al. "Integrated Methane Inversion (IMI 1.0): a user-friendly, cloud-based facility for inferring high-resolution methane emissions from TROPOMI satellite observations". In: *Geoscientific Model Development* 15.14 (2022), pp. 5787–5805. DOI: 10.5194/gmd-15-5787-2022. URL: https://gmd.copernicus.org/articles/15/5787/2022/.

[125] D. J. Varon et al. "Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes". In: *Atmospheric Measurement Techniques* 11.10 (2018), pp. 5673–5686. DOI: 10.5194/amt-11-5673-2018. URL: https://amt.copernicus.org/articles/11/5673/2018/.

[126] Daniel Varon et al. "Quantifying Time-Averaged Methane Emissions from Individual Coal Mine Vents with GHGSat-D Satellite Observations". In: *Environmental Science & Technology* XXXX (July 2020). DOI: 10.1021/acs.est.0c01213.

[127] Daniel Varon et al. "Satellite Discovery of Anomalously Large Methane Point Sources From Oil/Gas Production". In: *Geophysical Research Letters* 46 (Nov. 2019). DOI: 10.1029/2019GL083798.

[128] Daniel J. Varon et al. "Quantifying Time-Averaged Methane Emissions from Individual Coal Mine Vents with GHGSat-D Satellite Observations". In: *Environmental Science & Technology* 54.16 (2020). PMID: 32672947, pp. 10246–10253. DOI: 10.1021/acs.est.0c01213. eprint: https://doi.org/10.1021/acs.est.0c01213. URL: https://doi.org/10.1021/acs.est.0c01213.

[129] A. Vaughan et al. "CH4Net: a deep learning model for monitoring methane super-emitters with Sentinel-2 imagery". In: *EGUsphere* 2023 (2023), pp. 1–17. DOI: 10.5194/egusphere-2023-563. URL: https://egusphere.copernicus.org/preprints/2023/egusphere-2023-563/.

[130] A. T. Vermeulen et al. "COMET: a Lagrangian transport model for greenhouse gas emission estimation â€" forward model technique and performance for methane". In: *Atmospheric Chemistry and Physics Discussions* 6 (2006), pp. 8727–8779. DOI: 10.5194/acpd-6-8727-2006. URL: https://acp.copernicus.org/preprints/6/8727/2006/.

[131] Thumeera R Wanasinghe et al. "The internet of things in the oil and gas industry: a systematic review". In: *IEEE Internet of Things Journal* 7.9 (2020), pp. 8654–8673.

[132] J. Wang, A. P. Ravikumar, and A. R. Brandt. "Techno-economic Analysis of Deep-Learning-Enabled Automated Natural Gas Leakage Detection Technologies". In: *AGU Fall Meeting Abstracts*. Vol. 2019. Dec. 2019, GC51M-0960, GC51M–0960.

[133] J. Wang et al. "Deep Learning to Classify Methane Leak Size At Oil and Gas Facilities". In: 2018, A43R-3438 (Dec. 2018), A43R–3438.

[134] Jiayang Wang et al. "A Machine Learning Approach to Methane Emissions Mitigation in the Oil and Gas Industry". In: (Nov. 2020). DOI: 10.31223/X57W29.

[135] Jingfan Wang et al. "Machine vision for natural gas methane emissions detection using an infrared camera". In: *Applied Energy* 257 (2020), p. 113998. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2019.113998.

[136] Jingfan Wang et al. "VideoGasNet: Deep learning for natural gas methane leak classification using an infrared camera". In: *Energy* 238 (2022), p. 121516. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2021.121516. URL: https://www.sciencedirect.com/science/article/pii/S0360544221017643.

[137] Pin Wang, En Fan, and Peng Wang. "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning". In: *Pattern Recognition Letters* 141 (2021), pp. 61–67. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2020.07.042. URL: https://www.sciencedirect.com/science/article/pii/S0167865520302981.

[138] Shirui Wang et al. "Unsupervised Machine Learning framework for sensor placement optimization: analyzing methane leaks". In: (Jan. 2022).

[139] Yi Wang et al. "Remote sensing image super-resolution and object detection: Benchmark and state of the art". In: *Expert Systems with Applications* 197 (2022), p. 116793. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2022.116793. URL: https://www.sciencedirect.com/science/article/pii/S0957417422002524.

[140] Zhibao Wang et al. "An Oil Well Dataset Derived from Satellite-Based Remote Sensing". In: *Remote Sensing* 13.6 (2021). ISSN: 2072-4292. DOI: 10.3390/rs13061132. URL: https://www.mdpi.com/2072-4292/13/6/1132.

[141] Donald J Wuebbles and Katharine Hayhoe. "Atmospheric methane and global change". In: *Earth-Science Reviews* 57.3 (2002), pp. 177–210. ISSN: 0012-8252. DOI: https://doi.org/10.1016/S0012-8252(01)00062-9. URL: https://www.sciencedirect.com/science/article/pii/S0012825201000629.

[142] Haojun Xia, Alan Strayer, and Arvind P Ravikumar. "The Role of Emission Size Distribution on the Efficacy of New Technologies to Reduce Methane Emissions from the Oil and Gas Sector". In: *Environmental Science & Technology* 58.2 (2024), pp. 1088–1096.

[143] Yi-Jie Yang, Suman Singha, and Ron Goldman. "An automatic oil spill detection and early warning system in the Southeastern Mediterranean Sea". In: *EGU General Assembly Conference Abstracts*. EGU General Assembly Conference Abstracts. May 2022, EGU22-8408, EGU22–8408. DOI: 10.5194/egusphere-egu22-8408.

[144] Yi-Jie Yang, Suman Singha, and Roberto Mayerle. "A deep learning based oil spill detector using Sentinel-1 SAR imagery". In: *International Journal of Remote Sensing* 43.11 (2022), pp. 4287–4314. DOI: 10.1080/01431161.2022.2109445. eprint: https://doi.org/10.1080/01431161.2022.2109445. URL: https://doi.org/10.1080/01431161.2022.2109445.

[145] Xiao Youzi et al. "A review of object detection based on deep learning". In: *Multimedia Tools and Applications* 79 (Sept. 2020). DOI: 10.1007/s11042-020-08976-6.

[146] Daniel Zavala-Araiza et al. "A tale of two regions: methane emissions from oil and gas production in offshore/onshore Mexico". In: *Environmental Research Letters* 16 (Feb. 2021). DOI: 10.1088/1748-9326/abceeb.

[147] Haichao Zhang and Jianyu Wang. "Towards Adversarially Robust Object Detection". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Nov. 2019.

[148] Lu Zhang, Zhenwei Shi, and Jun Wu. "A Hierarchical Oil Tank Detector With Deep Surrounding Features for High-Resolution Optical Satellite Imagery". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.10 (2015), pp. 4895–4909. DOI: 10.1109/JSTARS.2015.2467377.

[149] Nannan Zhang et al. "Automatic Recognition of Oil Industry Facilities Based on Deep Learning". In: *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. 2018, pp. 2519–2522. DOI: 10.1109/IGARSS.2018.8518054.

[150] Siwei Zhang et al. "Atmospheric remote sensing for anthropogenic methane emissions: Applications and research opportunities". In: *Science of The Total Environment* 893 (2023), p. 164701. ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2023.164701. URL: https://www.sciencedirect.com/science/article/pii/S0048969723033247.

[151] Yifan Zhang, Junhui Hou, and Yixuan Yuan. "A Comprehensive Study of the Robustness for LiDAR-based 3D Object Detectors against Adversarial Attacks". In: (2023). arXiv: 2212.10230 [cs.CV].

[152] Yuzhong Zhang et al. "Quantifying methane emissions from the largest oil-producing basin in the United States from space". In: *Science Advances* 6.17 (2020), eaaz5120. DOI: 10.1126/sciadv.aaz5120. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.aaz5120. URL: https://www.science.org/doi/abs/10.1126/sciadv.aaz5120.

[153] Bryan Zhu et al. "METER-ML: A Multi-sensor Earth Observation Benchmark for Automated Methane Source Mapping". In: *arXiv preprint arXiv:2207.11166* (2022).

[154] Jiahe Zhu et al. "Transformer Based Remote Sensing Object Detection with Enhanced Multispectral Feature Extraction". In: *IEEE Geoscience and Remote Sensing Letters* (2023), pp. 1–1. DOI: 10.1109/LGRS.2023.3276052.

[155] Zijian Zhu et al. "Understanding the Robustness of 3D Object Detection with Bird's-Eye-View Representations in Autonomous Driving". In: CVPR '23. Mar. 2023.

[156] Daniel Zimmerle et al. "Methane Emissions from Gathering Compressor Stations in the U.S". In: *Environmental Science & Technology* XXXX (May 2020). DOI: 10.1021/acs.est.0c00516.

Résumé étendu (français)

Chapitre 1. Introduction

Le changement climatique est à l'origine de l'altération des modèles climatiques, y compris les augmentations des températures moyennes mondiales et les changements dans les schémas de précipitations. Le réchauffement climatique est un élément clé du changement climatique et est principalement causé par l'accumulation de gaz à effet de serre dans l'atmosphère terrestre, tels que le dioxyde de carbone et le méthane, dus aux activités humaines. Le méthane est responsable d'environ 50% [43] du réchauffement global depuis l'époque préindustrielle. Par définition, le méthane est un polluant climatique à courte durée de vie critique avec un potentiel de réchauffement global plus de 80 fois supérieur à celui du CO2 sur une échelle de temps de 20 ans. Le secteur Pétrolier et Gazier (P&G) est la deuxième plus grande source d'émissions anthropiques de méthane. Contrairement à l'agriculture, l'industrie P&G a une longue histoire de réduction des émissions de méthane en raison des préoccupations de sécurité, et le méthane peut être vendu comme gaz naturel, donc réduire les émissions a un avantage financier important. L'Agence internationale de l'énergie (AIE) estime que l'industrie peut réduire ses émissions mondiales de 78% et que jusqu'à 39% [41] de ces réductions peuvent être réalisées sans compromis financier.

Les stratégies efficaces de réduction du méthane, dépendent directement de la caractérisation précise des sources d'émission. Cependant, le manque de d'informations fiables copncernant les émissions de méthane a rendu difficile pour les gouvernements de mener des actions ciblées à l'échelle et à la vitesse nécessaires pour atteindre les objectifs de l'Engagement Méthane Mondial - Global Methane Pledge (GMP) -.

Motivations

Pour définir des objectifs, des réglementations et des stratégies spécifiques de reduction des emissions de méthane, les pays s'appuient sur les inventaires nationaux des émissions de méthane. Ces derniers sont principalement dérivés en utilisant des méthodes ascendantes qui estiment les émissions totales de méthane en utilisant des facteurs d'émissions. Cependant, diverses études (*e.g.* [1, 11, 93]) ont démontré la sous-estimation systématique des émissions de méthane de la chaîne d'approvisionnement P&G. Cette constatation remet en question l'utilisation des méthodologies ascendantes pour estimer les inventaires d'émissions, qui sont ensuite utilisés pour concevoir des lignes directrices réglementaires pour la réduction des émissions de méthane. Des informations incomplètes et non précises sur les niveaux réels d'émissions constituent un obstacle majeur à la réduction des émissions de méthane. Cependant, une quantité croissante de données sur les émissions de méthane émerge grâce au lancement régulier de nouveaux satellites dédiés à la mesure de la concentration de méthane avec une résolution spatiale de plus en plus élevée, une plus grande couverture et des seuils de détection plus sensibles. L'innovation technologique autour des outils de mesure du méthane en orbite tend à se concentrer sur la vision à long terme dans le domaine de la surveillance quasi-continue. Ces avancées dans les technologies de surveillance par satellite et leurs techniques de traitement sont un élément clé pour permettre la caractérisation du niveau et de la nature des émissions de méthane par une surveillance quasi-continue. En combinant les avantages respectifs des estimations ascendantes et descendantes des émissions de méthane, on améliorera finalement leur précision.

L'Observatoire International des Émissions de Méthane (IMEO) du Programme des Nations Unies pour l'Environnement (PNUE) a pour objectif de fournir des données en quasi en temps réel sur les émissions de méthane issues du secteur des combustibles fossiles. Il intègre des informations de diverses sources, y compris des satellites comme TROPOMI, GHGSat et MethaneSAT. Afin de pouvoir tirer partie de ces données et créer des informations utiles pour l'orientation des politiques de réduction des émissions de méthane, ces données d'émissions doivent être reliée à la source à leur origine pour caracteriser les comportements emetteurs de ces dernières. Pour ce faire, il est nécessaire d'attribuer toutes ces détections de méthane aux infrastructures, sites et opérateurs pétroliers et gaziers à leurs origines. Afin de parvenir à ces attributions pour chaque détection d'émissions de méthane de différents satellites à différents emplacements, il est nécessaire que le processus d'attribution repose sur des méthodes automatiques via

l'usage de l'intelligence artificielle.

Contributions

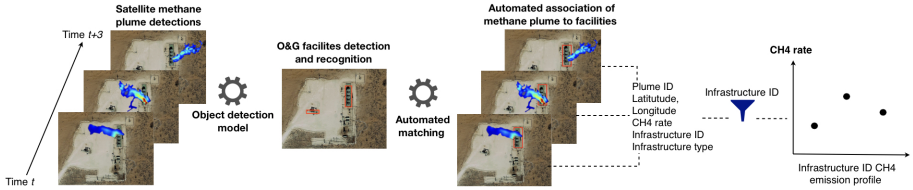


Figure 6.1: Méthode complète pour la détermination automatisée du profil d'émissions des infrastructures pétrolières et gazières. *Source des Image : @Google earth.*

- Conception d'un ensemble de données de référence d'images satellitaires à haute résolution annotée en fonction des infrastructures pétrolières et gazières;

- Détection et reconnaissance automatisées des infrastructures pétrolières et gazières basées sur des algorithmes de détection d'objets (fine-tuning et comparaisons de 3 types d'algorithme de detection d'objet);

- Détermination des effets du pré-entraînement des 3 algorithmes de détection d'objets sur leurs performances;

- Sensibilité et robustesse des 3 algorithmes et leurs modèles de détection d'objets aux attaques adverses basées sur les satellites;

- Association automatisée des émissions de méthane aux sites de la chaîne d'approvisionnement en pétrole et en gaz, aux opérateurs pétroliers et gaziers et à l'extension temporelle automatisée des campagnes de mesures passées;

- Méthode complète pour la détermination automatisée du profil d'émissions des infrastructures pétrolières et gazières illustrée par la Figure 6.1.

Chapitre 2. Émissions de méthane du secteur pétrolier et gazier

Le Chapitre 2 présente le contexte des émissions de méthane dans l'industrie pétrolière et gazière et son rôle dans le réchauffement climatique, les outils et méthodes existants pour mesurer les émissions de méthane ainsi que le cadre réglementaire actuel et ses défis.

Le méthane (CH_4) est un puissant gaz à effet de serre, dont l'impact sur le climat est estimé 84 fois supérieur à celui du dioxyde de carbone (CO_2) sur 20 ans et 28 fois sur 100 ans. Malgré sa courte durée de vie dans l'atmosphère, les concentrations de méthane ont presque triplé depuis l'ère préindustrielle, contribuant à environ 5% du réchauffement climatique. En 2022, les niveaux de méthane ont augmenté de 14,0 ppb, dépassant les prévisions, indiquant une tendance inquiétante et conduisant à une augmentation de 1,25°C des températures mondiales au cours du siècle dernier. Les émissions de méthane proviennent à la fois de sources naturelles et humaines. Les sources naturelles, principalement les zones humides, contribuent à environ 40 % des émissions mondiales de méthane, tandis que les sources anthropiques, notamment l'agriculture, la production d'énergie et la gestion des déchets, en représentent 60%. Ces émissions contribuent au réchauffement de la planète et peuvent entraîner la libération d'autres gaz à effet de serre provenant de sources telles que les hydrates de méthane présents dans le pergélisol et les océans. Les concentrations de méthane dans l'atmosphère sont mesurées en parties par million (ppm) ou en parties par milliard (ppb). Des capteurs sensibles au méthane, classés en deux catégories : in situ et à distance, sont utilisés pour quantifier les émissions. Les satellites sont des outils essentiels pour monitorer le méthane, chacun d'entre eux ayant des résolutions spectrales et spatiales uniques. L'estimation des émissions de méthane fait appel à deux approches : ascendante et descendante. Les méthodes ascendantes utilisent les données d'activité et les facteurs d'émission pour calculer les émissions, tandis que les méthodes descendantes utilisent les observations du méthane atmosphérique.

Le secteur des combustibles fossiles, y compris les activités pétrolières et gazières, a produit environ 135 millions de tonnes d'émissions de méthane en 2022, soit près de 40% des émissions de méthane d'origine humaine. Le secteur a un potentiel de réduction de 70% grâce aux technologies existantes, plus de 75% des émissions de méthane provenant d'activités en

amont telles que le forage et la production. Diverses politiques et initiatives ont été mises en place pour réduire les émissions de méthane, mais il n'existe pas de solution unique et l'efficacité des politiques dépend des circonstances spécifiques à chaque juridiction. L'estimation précise des émissions de méthane se heurte notamment à la sous-estimation systématique des émissions et à l'exclusion des super-émetteurs, qui représentent environ 40% des émissions totales de méthane. Des initiatives telles que le Partenariat pour le méthane dans l'industrie pétrolière et gazière (OGMP) et le Système d'alerte et de réponse pour le méthane (MARS) visent à fournir des données plus précises et plus complètes sur les émissions de méthane. La mise en œuvre de réglementations efficaces en matière d'émissions de méthane nécessite une nouvelle méthode d'inventaire fiable qui intègre des profils d'émissions de méthane à plusieurs niveaux. L'intelligence artificielle (IA) est essentielle pour traiter et analyser la grande quantité de données en temps réel et à l'échelle mondiale.

Chapitre 3. Détection et reconnaissance automatisées des infrastructures pétrolières et gazières

Le Chapitre 3 présente d'abord l'état de l'art des méthodes de monitoring des émissions de méthane, de la détection automatisée du panache de méthane à la prévision des émissions de méthane basée sur l'utilisation de l'intelligence artificielle. Il expose ensuite les performances de trois familles d'algorithmes de détection d'objets que nous avons adapté pour la reconnaissance automatique d'infrastructures pétrolières et gazières sur des images satellite à haute résolution. Il détaille également l'effet du préentraînement et des bruits de l'image satellite sur les performances de la détection d'objets.

Dans le contexte des émissions de méthane et de la surveillance des infrastructures P&G, les applications de l'IA se sont développées en raison de la disponibilité de vastes ensembles de données provenant de capteurs aériens et terrestres. Des techniques telles que les réseaux neuronaux convolutifs (CNN) sont utilisées pour la détection automatique des panaches de méthane, et l'IA est employée pour la prévision des émissions, la détection des défauts dans les pipelines, l'optimisation du placement des capteurs et la comparaison des méthodes traditionnelles de détection et de réparation

des fuites (LDAR) avec les méthodes LDAR basées sur la ML. Les algorithmes de reconnaissance d'objets, qui font partie de la vision par ordinateur, identifient et localisent automatiquement des objets dans des images ou des vidéos. Les méthodes basées sur les neurones sont généralement plus performantes que les techniques non neuronales et sont supervisées, nécessitant des bases de données d'images annotées pour l'apprentissage. Il existe différentes architectures de détection d'objets, classées en détecteurs à deux étapes, détecteurs à une étape, et d'autres comme DETR, un détecteur basé sur un transformateur. Dans les applications P&G, la détection d'objets est utilisée pour des problèmes de télédétection tels que la détection de déversements de pétrole, de réservoirs de pétrole, de puits, d'oléoducs et d'infrastructures P&G entières. Toutefois, le fait de se concentrer uniquement sur des infrastructures spécifiques pourrait ne pas rendre compte de la complexité des émissions de méthane dans le secteur de l'exploitation et de la gestion des ressources naturelles, d'où la nécessité d'adopter des approches plus larges incluant de multiples infrastructures.

Figure 6.2: Présentation des étapes de la méthode O&GProfile. Résultats de la détection visuelle d'objets par YOLO v8, FASTER-RCNN et DETR pré-entraînés sur 4 images test de la base de données OG (sources des images). @Google Earth).

Notre étude se concentre sur la réduction des émissions de méthane dans le

secteur du pétrole et du gaz en comprenant les profils d'émissions au niveau de l'opérateur individuel, du site et de l'infrastructure afin d'informer des mesures réglementaires efficaces. L'étude utilise le bassin permien aux États-Unis comme étude de cas, en employant et comparant trois algorithmes de détection d'objets : YOLO, FASTER-RCNN et DETR (cf. Figure 6.2) et les affine en utilisant une base de données spécialement conçue des infrastructures (P&G) avec des images satellite à haute résolution du bassin permien. L'étude évalue les algorithmes à l'aide de la précision moyenne (AP) et évalue leur robustesse par rapport aux variations des images satellites. Les principales conclusions de l'étude indiquent que YOLO v8 a dépassé les deux autres algorithmes en termes de précision dans la plupart des scénarios, que FASTER-RCNN a montré des performances supérieures dans l'identification de styles de compresseurs spécifiques, et que DETR a eu des performances adéquates mais a été généralement dépassé par YOLO v8. En termes de robustesse face aux variations des images satellite, FASTER-RCNN a été le moins affecté mais n'a pas réussi à détecter les puits, YOLO v8 a fait preuve d'une plus grande sensibilité mais a atteint la précision moyenne la plus élevée (mAP) après l'entraînement, et DETR a montré des performances variées face à différentes variations d'images mais a généralement été moins résilient que FASTER-RCNN. Nos expériences concluent que si YOLO v8 fait preuve de la plus grande précision dans la détection des infrastructures OG, le choix de l'algorithme doit tenir compte des exigences spécifiques et des types d'infrastructures à détecter. Il est également essentiel de prendre en compte l'impact des bruits communs que les images satellitaires peuvent contenir pour améliorer la robustesse des algorithmes dans les applications réelles. En effet, d'après nos expérimentations ces bruits tels que les distorsions tangentielles, les bruits Gaussien, etc. impactent tous d'une manière plus ou moins prononcé nos 3 algorithmes. Les directions de recherche futures suggérées par l'étude comprennent l'adaptation de la méthode à d'autres bassins de pétrole et de gaz, l'exploration d'autres paradigmes d'apprentissage tels que l'apprentissage "few-shot" et l'apprentissage "auto-supervisé", et l'étude de méthodes d'apprentissage d'ensemble pour combiner les points forts des algorithmes et modèles testés.

Chapitre 4. Attribution automatisée des panaches de méthane

Le Chapitre 4 présente une méthode d'association automatique des émissions de méthane aux différentes parties de la chaîne d'approvisionnement en pétrole et en gaz et aux opérateurs, basée sur l'utilisation de l'algorithme de clustering DBSCAN. Il décrit également notre méthode pour l'association automatique des panaches de méthane aux infrastructures pétrolières et gazières.

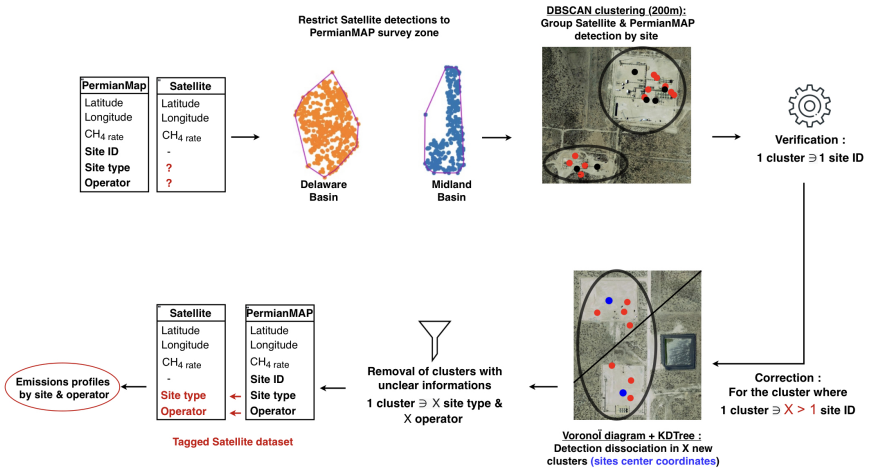


Figure 6.3: Présentation des étapes de la méthode O&GProfile.

Les progrès de la technologie des capteurs satellitaires point-source ont amélioré les estimations des émissions de méthane, permettant une surveillance quasi-constante sur de vastes zones. Cependant, le traitement de l'énorme quantité de données générées nécessite des techniques avancées de transformation des données, y compris des algorithmes d'apprentissage automatique. Les données des satellites spécialisés dans la détection du méthane fournissent des détails cruciaux sur chaque panache de méthane, mais l'identification de la source exacte d'une fuite de méthane nécessite l'intégration des données satellitaires avec des informations détaillées sur l'P&G au niveau du sol. Actuellement, il n'existe pas de méthodes automatisées permettant d'associer les détections de méthane à des sources spécifiques, ce qui rend le processus long et sujet aux erreurs. L'utilisation de l'IA et de la ML peut automatiser ce processus, améliorer l'efficacité et la

précision, et permettre une réponse plus dynamique aux fuites de méthane, conduisant à des actions opportunes pour atténuer les émissions et aider l'industrie de l'P&G à réduire son impact sur l'environnement.

Nos travaux sur l'attribution des sites et des opérateurs P&G à l'aide de données satellitaires se concentrent sur le bassin permien aux États-Unis, le plus grand bassin P&G des États-Unis, largement étudié pour les émissions de méthane. Notre étude utilise des données provenant du satellite Carbon-Mapper et des campagnes aériennes du GAO du projet PermianMAP lancé par l'Environmental Defense Fund (EDF). L'objectif principal est d'annoter les détections brutes provenant d'un satellite point-source "X" en utilisant les données PermianMAP afin de collecter des profils d'émission. La méthode proposée (cf. Figure 6.3), appelée P&GProfile, utilise l'algorithme de clustering DBSCAN pour regrouper les détections satellitaires et PermianMAP par site permettant de transferer les annotations (type de site et operateur) PermianMAP aux données satellitaires non annotées. Notre méthode utilise ensuite l'algorithme KDtree du plus proche voisin pour vérifier et corriger les clusters, ce qui permet d'obtenir un cluster pour chaque site. Le second volet de nos travaux se concentre également sur l'attribution des émissions de méthane à leurs infrastructures P&G grâce à des techniques avancées de détection d'objets, en utilisant des algorithmes de détection d'objets précédemment développés (cf. Chapitre 2.3) pour identifier et catégoriser les infrastructures P&G telles que les réservoirs, les puits et les compresseurs. Les panaches de méthane sont associés à l'infrastructure la plus proche en fonction de leur proximité spatiale, ce qui permet d'établir un lien précis entre les émissions de méthane et les infrastructures qui en sont la source, améliorant ainsi la surveillance et la compréhension des émissions de méthane provenant des activités P&G.

En conclusion, notre méthode O&GProfile emploie des techniques de regroupement pour l'attribution automatique des sites et des opérateurs des détections satellitaires d'P&G, atteignant un taux de réussite de 98,8% dans l'association des sites via DBSCAN, encore amélioré à 100% avec la correction Voronoi-NN-KDTree. Cette méthode permet de générer des profils d'émissions à partir des sites P&G, permettant une catégorisation par type de site et par opérateur, et peut être adaptée à différents ensembles de données et régions, bien qu'elle s'appuie sur des informations au sol qui peuvent devenir obsolètes en raison de l'évolution rapide du paysage P&G. La méthode suppose que l'infrastructure la plus proche de l'emplacement estimé d'une émission de méthane est la source d'émission, mais l'incertitude quant à l'emplacement de l'émission introduit une complexité dans ce processus d'attribution. Pour pallier ce problème, un périmètre d'incertitude

autour de l'emplacement estimé de chaque détection pourrait déterminer l'infrastructure source la plus probable, en exprimant les résultats de l'association sous forme de probabilités d'appartenance à des infrastructures proches.

Chapitre 5. Inventaire dynamique et intelligent des émissions de méthane

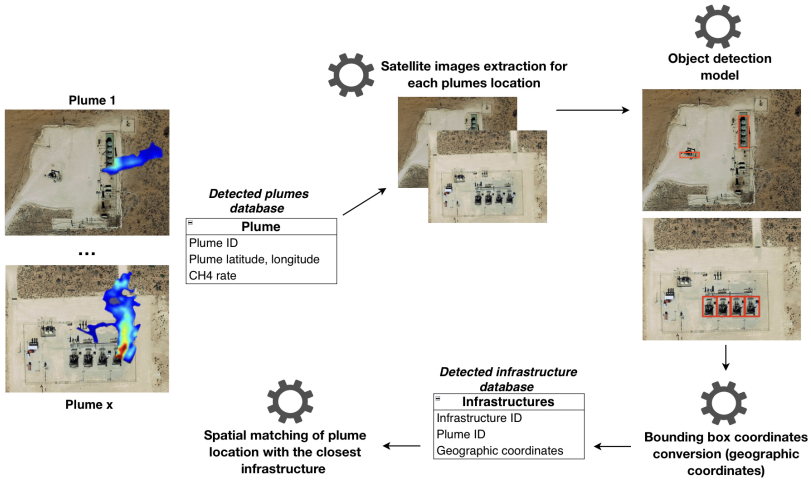


Figure 6.4: Présentation du cadre de l'inventaire dynamique du méthane. *Source : @Google earth.*

Le Chapitre 5 présente le cadre de l'Inventaire Dynamique et Intelligent des Émissions de Méthane (DIMEI), illustré par la figure 6.4, qui intègre des techniques de détection d'objets (Chapitre 2.3) et d'association automatique (Chapitre 2.4) pour améliorer la surveillance et l'inventaire des émissions de méthane des infrastructures pétrolières et gazières (P&G). Le cadre DIMEI est structuré de manière à permettre un monitoring en temps quasi-réel permettant la détermination des profils d'émissions aux niveaux des infrastructures, sites et opérateurs P&G. Ces derniers, collectés sur le long terme, pourront permettre de caractériser des événements spécifiques tels que les émissions volontaires et involontaires, de caractériser le comportement émetteur aux divers niveaux, mais aussi la prévision des spatio-temporelle des émissions de méthane. Le DIMEI se base sur base

de données de détections d'émissions de méthane par satellite, comprenant les coordonnées géographiques et les taux d'émission de ces derniers, qui sert de base aux processus analytiques du cadre. En effet, pour chaque panache de méthane détectés, le DIMEI utilise des algorithmes de détection d'objets pour identifier et classer les infrastructures P&G à partir d'images satellites à haute résolution (extraites à l'emplacement de chaque panache detecté) puis utilise un algorithme de clustering pour associer les panaches de méthane à leurs sources. L'aspect dynamique de DIMEI est illustré par sa capacité à collecter des informations temporelles et spatiales sur les émissions de méthane (cf. Figure ci-dessus), en générant des profils d'émission pour différents types d'infrastructures d'P&G au fil du temps.

L'aspect intelligent (à l'état de perspective) implique l'agrégation et l'analyse des profils d'émissions à différentes échelles, pouvant révéler les tendances, la saisonnalité et les variations irrégulières des émissions de méthane. Le potentiel d'intégration du DIMEI avec les réglementations existantes et futures en matière d'atténuation des émissions de méthane est détaillé, discutant sa capacité à informer l'élaboration des politiques, à faciliter le contrôle de la conformité et à améliorer l'efficacité de la réglementation. Le cadre DIMEI s'aligne sur des initiatives telles que l'OGMP et IMEO, visant à fournir des données précises et granulaires sur les émissions de méthane afin de soutenir les efforts mondiaux en matière de réduction du méthane. En conclusion, DIMEI représente un nouvel inventaire basé sur l'IA pour le suivi des émissions de méthane, offrant une approche dynamique et intelligente de la surveillance, de l'analyse et de la prévision des émissions. Bien qu'il soit actuellement axé sur le bassin permien et limité à certains types d'infrastructures P&G, la conception modulaire du cadre permet des améliorations et des extensions futures à d'autres régions et à d'autres types d'infrastructures. Des améliorations et des adaptations essentielles sont nécessaires pour permettre un fonctionnement entièrement automatisé, élargir le champ d'application des types d'infrastructures et améliorer la précision des données et la résolution temporelle afin d'obtenir des performances et une fiabilité optimales.

Chapitre 6. Conclusion

Dans cette thèse, nous avons présenté l'inventaire dynamique et intelligent des émissions de méthane (DIMEI), un nouveau système de surveillance des émissions de méthane pour trois types d'infrastructures P&G dans le bassin

permien, aux États-Unis, utilisant l'intelligence artificielle. Le DIMEI recueille de manière dynamique des données sur les émissions de méthane au fil du temps, fournissant des profils d'émissions qui caractérisent les comportements des différentes infrastructures, sites, bassins, pays et opérateurs de l'industrie pétrolière et gazière. Cette composante dynamique, exécuté sur le long terme, a pour objectif de fournir des series spatio-temporelles afin prévoir les émissions de méthane (aspect intelligent). En outre, l'aspect intelligent du DIMEI vise à créer une marge d'action pour réduire les émissions de méthane grâce aux prévisions, contribuant ainsi directement à la réduction du méthane et à l'élaboration de réglementations en la matière. Le DIMEI repose sur deux piliers principaux : la détection et la reconnaissance automatisées des infrastructures P&G dans le bassin permien et l'association automatisée des panaches de méthane détectés à des infrastructures P&G, des sites et des opérateurs spécifiques.

Nos contributions comprennent la mise au point méticuleuse et l'évaluation de trois algorithmes de détection d'objets conçus pour l'identification et la surveillance automatisées des infrastructures P&G dans le bassin permien. Nous avons développé une base de données inaugurale d'images d'infrastructures P&G grâce à l'annotation d'images satellites à haute résolution et avons effectué une analyse quantitative complète des impacts de la préformation sur les capacités de détection algorithmique. Notre nouvelle méthode a été introduite pour automatiser l'attribution des panaches de méthane détectés à des sites et opérateurs P&G spécifiques, reliant avec succès plus de 100 panaches de méthane à divers types de sites P&G et opérateurs dans le bassin permien. Notre méthode a indirectement étendu la couverture temporelle de l'étude PermianMAP et a fourni une méthode d'attribution automatisée des panaches de méthane détectés aux infrastructures P&G, tout en discutant des approches potentielles pour gérer les incertitudes associées à ces attributions. Le Chapitre 5 a présenté le premier DIMEI basé sur l'IA, combinant les méthodes des Chapitres 3 et 4, démontrant sa capacité à suivre de manière autonome les émissions de méthane dans le temps et détaillant les futures analyses automatisées potentielles basées sur les résultats DIMEI à moyen et long terme. Nous avons conclu en discutant de la symbiose entre DIMEI et les réglementations pour une réduction optimisée du méthane.

En ce qui concerne les travaux futurs, les perspectives à court terme comprennent l'amélioration des performances et de la robustesse de l'algorithme de détection des objets en élargissant la base de données et en utilisant des méthodes d'augmentation des données de télédétection, ainsi que l'amélioration de l'adaptabilité et de l'extension de DIMEI à d'autres infrastructures et

bassins de pétrole et de gaz par le biais de méthodes d'apprentissage supervisé ou de quelques clichés. En outre, la gestion des incertitudes liées au positionnement des panaches de méthane détectés est cruciale pour la rigueur du processus d'attribution. Les perspectives à long terme impliquent le développement de méthodes permettant de détecter et de reconnaître automatiquement les différents types de sites P&G et les exploitants, ce qui permet de s'affranchir de la dépendance à l'égard des études et des bases de données existantes qui deviennent rapidement inexactes en raison de l'évolution rapide du paysage P&G. En outre, l'automatisation de l'acquisition d'informations sur les sites des exploitants d'hydrocarbures à l'aide de méthodes de récupération sur le web à partir du programme Title V a été proposée comme une solution potentielle, bien qu'elle reste limitée. Pour le déploiement et l'utilisation futurs de DIMEI, des ajustements sont nécessaires pour intégrer diverses sources satellitaires fournissant des estimations différentes et pour utiliser une API pour l'extraction automatique et optimisée dans le temps d'images satellitaires sur les sites de détection du méthane. DIMEI générant de grandes quantités de données sur les émissions, de nouvelles méthodes d'élaboration des politiques basés sur l'intelligence artificielle seront nécessaires pour traiter efficacement ces données et les transformer en informations exploitables pour concevoir des réglementations sur l'atténuation du méthane, ce qui nécessitera des changements dans les méthodes réglementaires pour s'adapter à ce nouveau type d'inventaire.